University of Essex

The role of cognitive individual differences and learning difficulty in instructed adults' explicit and implicit knowledge of selected L2 grammar points: A study with Mexican learners of English

Luis Humberto Rodríguez Silva

A thesis submitted for the degree of Doctor of Philosophy in

Applied Linguistics

Department of Language and Linguistics

University of Essex, United Kingdom

November, 2017

Supervisor: Dr. Karen Roehr-Brackin

Table of Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor Dr. Karen Roehr-Brackin who always supported me in the process of writing this thesis. The completion of my thesis has only been possible to your patience, knowledge, and academic support. I'm grateful because you kept my motivation on my research topic from beginning to end. I'm also indebted to you for all your support for the conferences I presented on my research topic at University of Essex, University of Cambridge and my home university Universidad Autónoma de Aguascalientes, and especially for your invitation to co-author an article in the journal *Studies in Second Language Acquisition* and the presentation at University of Basel in Switzerland. I am very happy you were my supervisor all these years. I knew from your first classes and research feedback sessions that you were the one indicated to supervise my work, and that the Department of Language and Linguistics at the University of Essex has an invaluable asset in both the academic and research area. You have my heartfelt gratitude for your feedback on every word, line, section, chapter of my thesis and the thesis as a whole. Thank you very, very much, Karen!

Secondly, I am indebted to the Second Language Research Group in the Department of Language and Linguistics at the University of Essex for their feedback on the instruments on this study. I am also indebted to the teachers and students who took part in my research; I have no words to thank you enough for your time and willingness that you showed in every possible way during the administration of the instruments. Thank you very much to you all!

Thirdly, I am also indebted to the authorities in my university in Mexico, Universidad Autónoma de Aguascalientes, for the academic leave they granted me to undertake the program of studies of Ph.D. (Integrated) in Applied Linguistics at the University of Essex in England.

Fourthly, I would not have completed my Ph.D. study journey without the financial support from the Mexican government; they supported me with the grant no. PROMEP/103.5/11/5527 from the Programa de Mejoramiento del Profesorado (Teaching Profession Improvement Program) which is a Federal Government program in Mexico administrated by the Sub-Ministry of Higher Education. Thank you for sponsoring me for such an endeavour!

Finally, I especially dedicate this thesis to my beloved wife and children in recognition of their constant support and encouragement. Thank you all very, very much!

The role of cognitive individual differences and learning difficulty in instructed adults' explicit and implicit knowledge of selected L2 grammar points: A study with Mexican learners of English

ABSTRACT

This study explored the relationship between implicit and explicit knowledge of 13 second language (L2) English grammar points and the relationship of each type of knowledge with language learning aptitude and working memory capacity in 90 Mexican learners of L2 English at three different levels of proficiency (Level 5, Level 7, Level 9). An elicited imitation test and an oral narrative test were used to measure implicit knowledge, and a metalinguistic knowledge test was used to measure explicit knowledge. With respect to language learning aptitude and working memory, the former was operationalised by the LLAMA test, and the latter by the backward digit span test. With regard to the relationship between implicit and explicit knowledge, implicit knowledge correlated positively and weakly with explicit knowledge while an analysis by grammar point showed a non-significant negative correlation approaching significance between implicit and explicit scores. These results indicate that learners found some grammar points easy in terms of explicit knowledge and other grammar points easy in terms of implicit knowledge, and vice versa. Learners' language aptitude and working memory did not significantly predict explicit or implicit knowledge of the targeted difficult and easy grammar points for the cohort of participants as a whole. Another analysis by level group (Level 5, Level 7, Level 9) indicated that the cognitive variables did not significantly predict explicit knowledge of easy or difficult grammar points. However, with respect to implicit knowledge, working memory significantly predicted implicit knowledge of easy grammar points in Level 5, and language aptitude marginally predicted implicit knowledge of difficult grammar points in Level 5. Overall, the findings support the view that language aptitude and working memory are better predictors at lower levels of proficiency. The findings of this study contribute to researchers'

understanding of the distinction between implicit and explicit knowledge and the relevance of language aptitude and working memory at different levels of proficiency.

CHAPTER ONE: INTRODUCTION

Learning a foreign language is an evolving and dynamic process that does not occur instantly or that requires just a few language lessons (N. Ellis, 2011; Spada & Lightbown, 2008); it requires a great deal of practice. DeKeyser (1998: 50) uses the term practice "in the sense of engaging in an activity with the goal of becoming better at it" (p. 50). He argues that some teaching methods and approaches do not implement the practice of a second language (L2) in the way he defines it; this implies that some teachers may focus more on form than on meaning and others more on meaning than on form when a balance should be kept between form, meaning, and use.

In an early study, Sorace (1985) investigated the relationship between knowledge and use of an L2 (Italian) in L1 English speakers. She concluded that despite learners' exposure to a grammar-oriented teaching method, they have difficulty in expressing themselves in communicative situations. This type of situation has led to a change from grammar-based instruction to a "grammar-free" instruction and back to the inclusion of explicit instruction (N. Ellis, 2011) due to the difficulties learners experience with pronunciation, morphological, syntactic, and pragmatic features of the L2 in methods such as Communicative Language Teaching and Content Based Instruction (Spada & Lightbown, 2008) where little or no emphasis is paid to the explicit learning of grammar points. Hence, instruction focused exclusively on form or exclusively on meaning may not be as effective as instruction focused on both form and meaning for developing both explicit and implicit knowledge of the target language. In this sense, explicit instruction that encourages learners to use the language may help learners improve their ability to communicate fluently and confidently in an L2, which "is considered to be the ultimate goal of instruction" (Akakura, 2012, p. 10).

Nevertheless, developing explicit and implicit knowledge, a key component to improve learners' ability to communicate fluently and confidently in an L2, involves several factors such as the difficulty of grammar points, and learners' individual differences (IDs). With regard to the difficulty of grammar points, how to determine which grammatical structures are easy or difficult to learn in terms of implicit and explicit knowledge can be done following various approaches (Collins, Trofimovich, White, Cardoso, & Horst, 2009; R. Ellis, 2006; Spada & Tomita, 2010; Roehr & Gánem-Gutiérrez, 2009a), and it appears that the administration of a difficulty judgement questionnaire to both teachers and learners is a reliable measure (Absi, 2014; Rodríguez Silva & Roehr-Brackin, 2016; Scheffler, 2011).

Concerning IDs, it is known that the emphasis towards understanding how L2 learners learn has often been placed on the average speaker or listener without paying much attention to IDs (Roberts & Meyer, 2012), in particular to language learning aptitude and working memory (WM) capacity, which are related to the learning difficulty of grammar points. In other words, whatever difficulty learners have in learning L2 grammar points, this is not only due to the complexity of the structures themselves but also to their IDs (Bulté & Housen, 2012; Dörnyei, 2009; Housen & Simoens, 2016; Robinson, 2002). DeKeyser (2003) refers to the latter as subjective difficulty when it comes to learning grammar points, namely, he points out that "subjective difficulty refers to the actual difficulty that individual learners experience when learning a second language" (p. 431).

With respect to the factor language aptitude, traditional aptitude tests were developed to predict rate and success of foreign language learning (Dörnyei, 2005) and areas of difficulty in early stages of second language acquisition (SLA) (Robinson, 2005). On the other hand, contemporary aptitude research has focused on predicting high levels of attainment and how learners can benefit from incidental exposure to the L2 (Robinson, 2005). A recent development in aptitude research is the concept of aptitude-structure interaction (Erlam, 2005; DeKeyser, 2012), that is, language aptitude in relation to different types of rules and uses. Particularly, little research has been conducted on the interaction of language learning aptitude and implicit/explicit knowledge of grammar points. This is an important issue because language aptitude may play a role for structures with high learning difficulty but not for structures with low learning difficulty. Similarly, little research has been conducted on the interaction of working memory and implicit/explicit knowledge of grammar points; working memory may also play a role for structures with high learning difficulty but not for structures with low learning difficulty.

Regarding the implicit and explicit knowledge of an L2, a number of studies have used a variety of implicit and explicit measures to gauge implicit and explicit knowledge of instructed learners (Absi, 2014; Ellis, 2006; Erlam, 2006; Rodríguez Silva & Roehr-Brackin, 2016) showing that it is possible to measure both types of knowledge to some extent.

Research studies have also been conducted on the relationship between implicit and explicit knowledge of a number of specific grammar points (Bowles, 2011; R. Ellis, 2006; Han & Ellis, 1998; Rodríguez Silva & Roehr-Brackin, 2016), between implicit and explicit knowledge and language aptitude (Yalçin & Spada, 2016), between implicit and explicit knowledge and WM (Erçetin & Alptekin, 2013; Serafini & Sanz, 2015), and between implicit and explicit knowledge, language aptitude and WM (Erlam, 2005).

Overall, the studies involving implicit and explicit knowledge showed that participants performed better on the explicit measures than the implicit measures. Furthermore, correlational analyses revealed weak and statistically non-significant correlations between implicit and explicit scores (Han & R. Ellis, 1998; R. Ellis, 2006) as well as weak to moderate statistically significant correlations (Bowles, 2011; Rodríguez Silva & Roehr-Brackin, 2016). Likewise, the studies involving implicit/explicit knowledge, language aptitude and working memory revealed that language aptitude correlated with easy and difficult grammar points (Yalçin & Spada, 2016), and working memory correlated with implicit and explicit L2 knowledge (Erçetin & Alptekin, 2013; Serafini & Sanz, 2015; Yalçin & Spada, 2016). Erlam's (2005) study seems to be the only study that included implicit and explicit L2 knowledge, language aptitude and WM as well as three instructional methods (deductive instruction group, inductive instruction group, structured input instruction group). The findings indicate that different methods benefit differentially different learners with different profiles, that is, learners with high language aptitude or working memory may benefit more from one type of instruction than the others.

While a number of studies have been conducted on explicit knowledge (Alderson, Clapham, & Steel, 1997; Bialystok, 1979; Sorace, 1985, Ziętek & Roehr, 2011) implicit knowledge (Erlam, 2006), and on both explicit and implicit knowledge (Ellis, 2006) on a number of grammatical structures, no studies have been conducted on the interaction between individual differences variables (language learning aptitude and working memory) and linguistic variables (implicit and explicit knowledge of grammar points) with learners at different levels of proficiency and focusing on a range of grammar points that vary in terms of learning difficulty. In this context, the present study addresses the aforementioned gaps in applied linguistics research and SLA by a) categorizing 13 grammar points according to teachers' and learners' perceived difficulty, b) recruiting student participants from three different levels of proficiency, c) incorporating both implicit and explicit L2 measures, and d) examining separately the relationship of complex WM and implicit and explicit L2 knowledge of the targeted grammar points, the relationship of language aptitude and implicit and explicit L2 knowledge of the targeted grammar points, and the relationship of complex WM and language aptitude.

This research topic on language learning aptitude, working memory and their relationship to explicit and implicit L2 knowledge is relevant not only for the field of SLA but also for the English language teaching profession which often treats learners as average speakers or listeners (Roberts & Meyer, 2012; Skehan, 2012) without taking into account that foreign language aptitude and working memory as individual difference variables can influence foreign/second language learning in formal settings with adolescents and adults (Goo, Granena, Yilmaz & Novella, 2015; Ortega, 2009) in both implicit and explicit conditions (Dörnyei, 2005).

This thesis is organized as follows. Chapter 2 introduces the main variables of implicit and explicit knowledge, learning difficulty of grammar points, language learning aptitude, and working memory capacity and provides definitions for each concept as well as the corresponding theoretical background. Chapter 2 also presents studies related to the different variables and provides an overview of associations between implicit and explicit knowledge, between implicit/explicit knowledge and language aptitude, and between implicit/explicit knowledge and working memory. The chapter discusses the interface hypothesis between implicit and explicit knowledge, particularly, the possibility that explicit instruction may impact on both implicit and explicit learning and knowledge. In the last section of Chapter 2 the research questions are presented. Chapter 3 presents the main study procedures, the administration of implicit and explicit measures, the measure of language aptitude, and the measure of working memory. It also presents the difficulty judgement of the 13 grammar points made by student and teacher participants. In Chapter 4, the results are presented in detail commensurate with the research questions, and these are interpreted and discussed in relation to previous findings in Chapter 5. Finally, Chapter 6 sums up the main findings of the study, highlights its limitations, and concludes with suggestions for further research.

CHAPTER TWO: LITERATURE REVIEW

This chapter is divided into four sections. The first section begins with the introduction to the different types of learning in second language acquisition and the definitions of learning, implicit and explicit learning, implicit and explicit knowledge, and metalinguistic knowledge. Then, the key characteristics of implicit and explicit knowledge are presented. Linked to these key characteristics, arguments on the interface positions between implicit and explicit knowledge are discussed as well as the operationalisation of these two constructs. In addition to this, empirical findings of implicit and explicit knowledge and the association between these two constructs are presented. In the second section, the relative difficulty of grammar points pertaining to implicit and explicit L2 knowledge is addressed. This relative difficulty is depicted in terms of a structural complexity and cognitive difficulty perspective. A number of approaches to determine learning difficulty of grammar points are discussed such as taxonomies of implicit and explicit learning difficulty of grammar points (Ellis, 2006; DeKeyser, 2005; Roehr & Gánem-Gutiérrez, 2009), the number of criteria applied to arrive at the correct target form (Hulstijn & de Graaff, 1994, Spada & Tomita, 2010), the expert judgement of experienced L2 teachers (Robinson, 1996; Scheffler, 2011), and the combined judgements of both experienced L2 teachers and learners of different levels of proficiency (Absi, 2014; Rodríguez Silva & Roehr-Brackin, 2016). In the third section, language learning aptitude is presented as an influential predictor of learners' performance on measures of implicit and explicit knowledge. Particularly, language aptitude as operationalised by the LLAMA aptitude test is a plausible option for testing learners' language aptitude in instructional settings. In the fourth section, working memory (WM) is presented as another important individual differences (IDs) factor in the language learning process, and like language aptitude, it is also influential in learners' performance on measures of implicit and explicit knowledge. A number of WM models are presented singling out Baddeley and Hitch's (1974) model as one of the most influential working memory models in SLA research. A distinction between measures of complex WM and phonological short-term memory is made, and the relation between WM capacity and language proficiency, WM capacity and language aptitude are discussed. In the last two sections of this chapter, a summary of studies conducted on the variables mentioned above is presented first and then the research questions that will be addressed in the main study are outlined.

2.1 Implicit and explicit knowledge and learning in SLA

It is undisputable that individuals show more heterogeneity when they learn a second language (L2) (DeKeyser, 2005; Granena, 2014; Hummel, 2009; Kormos & Sáfár, 2008; Ortega, 2009; Roberts & Meyer, 2012) than when they learn their mother tongue (L1) (Hulstijn, 2005; Rubin, 1975; Sawyer & Ranta, 2001). The relative invariability in L1 learning outcomes is mainly due to the implicit processes children use. This invariability is evidence, in most cases, of the children's development of their full communicative competence of their L1, with the exception of those who may not be exposed to sufficient quantities of input and those who may suffer from impairments (Hulstijn, 2005). In contrast, the variability in L2 learning outcomes among adult learners is at least in part attributable to differences in their analytic capacity which is reflected in their explicit learning, though it is undeniable that L2 learning also involves implicit learning. Hence, learning an L2 involves both implicit and explicit learning (whether it is intentional, incidental, inductive or deductive) (Akakura, 2012; N. Ellis, 2011; R. Ellis, 2006; Erlam, 2005).

2.1.1 Implicit and explicit learning

When dealing with the notion of learning, it is crucial to firstly understand its meaning before advancing to the concepts of implicit and explicit learning. Richards and Schmidt (2002) define learning as "the process by which change in behaviour, knowledge, skills, etc., comes about through practice, instruction or experience" (p. 298). This general process is represented by both implicit and explicit learning when it comes to the acquisition of an L2 (DeKeyser, 2003; N.

Ellis, 2005; R. Ellis, 2004; Krashen, 1981). What follows is the presentation of two different definitions of implicit and explicit learning by Arthur Reber, a cognitive psychologist who was interested in learning in more general terms, and by Jan Hulstijn, an applied linguist who has shown specific interest in language learning. Reber (1976), the pioneer of implicit learning research, defined implicit learning as “a primitive process of apprehending structure by attending to frequency cues” and explicit learning as “a more explicit process whereby various mnemonics, heuristics, and strategies are engaged to induce a representational system” (p. 93). These two definitions show the presence or absence of conscious intention to learn something. The following two recent definitions which are of theoretical and practical relevance elucidate such a distinction; Hulstijn (2005) defines explicit learning as “input processing with the conscious intention to find out whether the input information contains regularities and, if so, to work out the concepts and rules with which these regularities can be captured” and implicit learning as “input processing without such an intention, taking place unconsciously” (p. 131). Whereas Reber (1976) employs the key terms of primitive process and explicit process, and Hulstijn (2005) conscious intention and no intention to distinguish between implicit and explicit learning, both describe these two types of learning process by stating that the implicit process is unconscious and the explicit process is conscious (DeKeyser, 1995; N. Ellis, 2011; R. Ellis, 2005). These two notions comprise the internal processes learners go through to learn an L2, and the result of each type of process is implicit and explicit knowledge.

2.1.2 Implicit and explicit knowledge

Like the variety of definitions existing in the research literature on the issue of implicit and explicit learning, a number of definitions have been proposed on the topic of implicit and explicit knowledge. The latter is understood as a synonym of metalinguistic knowledge (Bialystok, 1994; Han & Ellis, 1998; R. Ellis, 2004). In the most general terms, metalinguistic knowledge has been defined as learners’ explicit knowledge about language (Alderson et al., 1997; Bialystok, 1979;

Elder, Warren, Hajek, Manwaring, & Davies, 1999; Ellis, 2004). In specific terms, Roehr (2008) drawing on Hu (2002) and R. Ellis (2004), defines L2 metalinguistic knowledge as “a learner’s explicit knowledge about the syntactic, morphological, lexical, phonological, and pragmatic features of the L2. It includes explicit knowledge about categories as well as explicit knowledge about relations between categories” (p. 179). Anderson (2005) and Hulstijn (2005) define explicit knowledge as declarative knowledge that can be brought into awareness and that is potentially available for verbal report, and implicit knowledge as knowledge that cannot be brought into awareness or articulated. In this sense, both implicit and explicit knowledge are represented in learners’ mental processes when they engage in learning grammatical structures either as implicit or explicit knowledge.

So far, the distinction between the notions of implicit and explicit learning and implicit and explicit knowledge has been introduced through their definitions. R. Ellis (2004) has proposed a list of characteristics of implicit and explicit knowledge (see Table 2.1), and subsequently, a checklist of criteria to help operationalise these two constructs in order to design tests to measure them (see Table 2.2).

Table 2.1 Key characteristics of implicit and explicit knowledge

Characteristics	Implicit knowledge	Explicit knowledge
Awareness	Intuitive awareness of linguistic norms	Conscious awareness of linguistic norms
Type of knowledge	Procedural knowledge of rules and fragments	Declarative knowledge of grammatical rules and fragments
Systematicity	Variable but systematic knowledge	Anomalous and inconsistent knowledge
Accessibility	Access to knowledge by means of automatic processing	Access to knowledge by means of controlled processing
Use of L2 knowledge	Access to knowledge during fluent performance	Access to knowledge during planning difficulty

Self-report	Nonverbalisable	Verbalisable
Learnability	Potentially only at an early age	At any age

(Adapted from R. Ellis, 2005, p. 151)

2.1.3 Operationalisation of implicit and explicit knowledge

According to R. Ellis (2005), Table 2.2 informs the design of measures of implicit and explicit knowledge. Particularly, the criteria systematicity and certainty inform what is expected learners will respond on each type of measure.

Table 2.2 Operationalising the constructs of L2 implicit and explicit knowledge

Criterion	**Implicit knowledge**	**Explicit knowledge**
Degree of awareness	Response according to feel	Response using rules
Time available	Time pressure	No time pressure
Focus of attention	Primary focus on meaning	Primary focus on form
Systematicity	Consistent responses	Variable responses
Certainty	High degree of certainty in responses	Low degree of certainty in responses
Metalinguistic knowledge	Metalinguistic knowledge not required	Metalinguistic knowledge encouraged

(Adapted from R. Ellis, 2005, p. 152)

Table 2.2 shows a clear distinction between the operationalisation of implicit and explicit knowledge, which can be interpreted as the former being more intuitive and the latter more controlled when a learner is engaged in language use.

However, it is worth noting that despite the consensus that seems to exist among SLA researchers about time pressure as one of the main factors for accessing implicit knowledge, there is the possibility that even under time pressure participants may use automatised explicit knowledge (Susuki & DeKeyser, 2015). Similarly, on measures of explicit knowledge such as untimed grammaticality judgement tests and metalinguistic knowledge (MLK) tests, there is the possibility that participants may use implicit knowledge on sections requiring error correction of grammar points. Despite these issues, researchers (Absi, 2014; R. Ellis, 2005; Rodríguez Silva & Roehr-Brackin, 2016; Roehr & Gánem-Gutiérrez, 2009a; Scheffler, 2011) have operationalised implicit and explicit knowledge with some success. For instance, Ellis' (2005) operationalisation of the constructs of implicit and explicit L2 knowledge allows designing tests either for teaching or researching purposes (Gutiérrez, 2013; Hulstijn, 2005).

2.1.4 Measures of implicit and explicit knowledge

A number of studies has employed measures for either type of knowledge or has included measures for both types of knowledge (Absi, 2014; Akakura, 2012; R. Ellis, 2006; Erlam, 2006; Norris & Ortega, 2001; Rodríguez Silva & Roehr-Brackin, 2016; Scheffler, 2011; Spada & Tomita, 2010; Ziętek & Roehr, 2011). In meta-analyses, the range of inclusion of measures of implicit and explicit knowledge has shifted from the use of fewer implicit measures (Norris & Ortega, 2001) to more implicit measures (Goo et al., 2015; Spada & Tomita, 2010). In the former meta-analysis, only 10% of the studies used free outcome measures whereas in the latter meta-analyses 17% and 33.3% of the studies used free outcome measures, respectively. Among the measures of implicit knowledge that are widely used in SLA research are elicited imitation tests, timed grammaticality judgement tests, oral narrative tests, and speaking tests based on description of pictures, discussion of topics, and report of plans or activities, and the measures of explicit knowledge are typically untimed grammaticality judgement tests and metalinguistic knowledge tests (Absi, 2014; Norris & Ortega, 2001; Spada & Tomita, 2010).

With regard to the implicit measures (i.e. the elicited imitation test and the timed grammaticality judgement test) in R. Ellis' (2006) study, a total of over 220 L2 learners of different levels of language proficiency and mixed L1s took these two tests. The elicited imitation test (EI test) consisted of a set of 34 belief statements targeting 17 grammatical structures. The 34 statements (one grammatical and one ungrammatical sentence per structure) were presented aurally to participants. The participants were required to say first whether they agreed with, disagreed with or were not sure about each statement and then to repeat the sentences orally in correct English. Their responses were audio-recorded. The results indicate that participants found this test difficult (mean % = 50.44, SD = 18.91). On the other hand, the timed grammaticality judgement test was a computer-delivered test consisting of 68 sentences, evenly divided between grammatical and ungrammatical. The sentences, which were different from those in the imitation test, were presented in written form on a computer screen. The time allowed for judging the individual sentences ranged from 1.8 to 6.24 seconds. There were four sentences to be judged for each of the 17 grammatical structures. The results indicate that participants found this test difficult as well (mean % = 56.21, SD = 11.88) but somewhat easier than the EI test.

In the same study (R. Ellis, 2006), the untimed grammaticality judgement test and MLK test comprised 68 sentences, evenly divided between grammatical and ungrammatical. The sentences were computer-delivered in written form. Participants were required to indicate in their own time whether each sentence was grammatical or ungrammatical. The MLK test consisted of two parts, but only the scores from part 1 were reported in the study. This part of the test presented participants with seventeen ungrammatical sentences, based on the seventeen structures, and required participants to select the rule that best explained each error out of four choices provided. The result obtained for the untimed GJT was much higher (mean % = 80.67, SD = 13.13) than the one obtained for the timed GJT (mean % = 56.21, SD = 11.88), and similarly, the result obtained for the MLK test (mean % = 54.61%, SD = 15.56) was slightly higher than the result obtained for the oral imitation test (mean % = 50.44, SD = 18.91).

Furthermore, R. Ellis (2006) conducted two factor analyses to investigate the extent to which the oral imitation test and the timed GJT measured implicit knowledge and the untimed GJT and

MLK test explicit knowledge. In the first analysis the total scores for the four tests were entered; the results showed that the oral imitation test and the timed GJT loaded on factor 1 (implicit knowledge) and the MLK test loaded on factor 2 (explicit knowledge) while the untimed GJT loaded on both factors. The researcher carried out a second analysis, substituting the untimed GJT (ungrammatical sentences) scores for the untimed GJT (total) scores. The results of this second analysis showed that the oral imitation test and the timed GJT loaded on factor 1, and the MLK test and the untimed GJT loaded on factor 2. These analyses are presented as evidence that these tests are measures of implicit and explicit knowledge respectively. Furthermore, the construct validity of the battery of tests in R. Ellis (2006) has been supported by the results obtained in Bowles (2011), Gutiérrez (2013), and Zhang (2015) in different learning environments, and with learners from different L1 backgrounds.

With regard to grammaticality judgment tests and MLK tests, Gutiérrez (2013) examined the construct validity of these measures. The study involved 49 participants at a Canadian university. Twenty-nine of the participants were near the end of their third term of Spanish language instruction (A2 level of the Common European Framework of Reference (CEFR))[1]. The other 20 participants were close to completing their fifth term of Spanish language instruction (B1 level). The participants completed a timed grammaticality judgement test, an untimed grammaticality judgement test, and a MLK test. The tests were designed following the criteria in R. Ellis (2005). Both grammaticality judgement tests contained 64 sentences, half of which were grammatical and half of which were ungrammatical. The 64 sentences covered 16 grammatical structures. For each of the grammatical structures, there were two grammatical sentences and two ungrammatical ones. The sentences were the same in both the timed and untimed grammaticality judgement test. For the timed grammaticality judgement test, the sentences were presented on an automated PowerPoint slide show, and the participants were asked to write their responses on an

[1] In the CEFR, A1 is the lowest level, whereas C2 is the highest.

answer sheet. For the untimed grammaticality judgement test, the participants received an answer sheet and a separate sheet containing the sentences, and were instructed to indicate whether the sentences were grammatical or ungrammatical on the answer sheet; there was no time limit to complete this test. The MLK test consisted of 16 sentences covering the same grammatical structures as the grammaticality judgement tests. Each sentence contained an underlined error, and the participants were asked to provide a written explanation in English of the rule violated in each of the sentences; there was no time limit to complete this test. The grammatical and ungrammatical sections in both the timed and untimed grammaticality judgement tests correlated significantly with the MLK test scores, although the correlations with the ungrammatical sections were stronger than the ones with the grammatical sections. A principal component factor analysis showed that the ungrammatical sections of both timed and untimed grammaticality judgement tests and the MLK test loaded on Factor 1 (i.e. explicit knowledge), whereas the grammatical sections of both tests loaded on Factor 2 (i.e. implicit knowledge). If a comparison is made between the loadings of the factor analyses of R. Ellis (2006) and Gutiérrez (2013) regarding the timed/untimed and grammatical/ungrammatical grammaticality judgement test items, it is not entirely clear whether one type of test or the other measures implicit or explicit knowledge.

In Ziętek and Roehr (2011), a MLK test was aimed at assessing the participants' level of explicit knowledge about the English language. The participants were 20 second-year students at a public lyceum in Wroclaw, Poland. A lyceum is the most common type of secondary school in Poland. The test consisted of 24 sentences targeting twelve grammar points of English. The 24 sentences contained one highlighted error each. Participants were required to correct, describe, and explain the highlighted errors, and they could use their L1 (Polish) or English to complete the test. An example is provided below.

Example:

Sentence: If I have had enough money last year, I would have bought a house.

Correction: ________________________

Description/explanation:

____

____

Following most of the criteria in R. Ellis' (2005) operationalisation of the constructs of L2 implicit and explicit knowledge (i.e., degree of awareness, time available, focus of attention, systematicity, certainty, metalinguistic knowledge—see Table 2.2), Ziętek and Roehr's (2011) MLK test required test takers (a) to focus on form because they have to describe and explain the grammar point highlighted in the sentence, that is, in their description they had to specify what form and in their explanation they had to explain why that form; (b) they were given enough time to allow access to their explicit knowledge of the selected grammar points and, (c) the degree of awareness that was required was high because of the verbalization required. The results indicate that participants' overall performance on the MLK was somewhat low (mean % = 54). The MLK correction section resulted in a much higher score (mean % = 91) than that of the description/explanation section (mean % = 34).

In another study (Scheffler, 2011), 50 Polish secondary learners of English at the upper intermediate level were asked to complete a MLK test consisting of 12 rules which were taken from the books that the learners had used and were using at the time of the study. The learners received instructions in Polish to illustrate in writing as accurately as possible each of the rules in the test with one sentence in English. Like Ziętek and Roehr's (2011) test, the same four criteria (i.e., degree of awareness, time available, focus of attention, metalinguistic knowledge) apply to this test. An example is provided below.

Example:

Rule: We use the present simple to talk about habits and routines.

Example sentence: ____

The results indicate that learners' metalinguistic knowledge on the 12 rules varied considerably. The overall success rate was 71.6%. Unlike the previous MLK test, this test precludes the correction section leaving the only possibility for learners to access their metalinguistic knowledge to write the example sentence in accordance to each grammatical rule.

With respect to speaking tests, in Absi's (2014) study, 64 Syrian undergraduate students studying English literature at university level completed a speaking test consisting of 14 oral elicitation tasks. These tasks were planned to elicit prototypical and peripheral uses of ten grammar points. Most of the tasks involved the use of information cards and photos. The participants were required either to describe or compare between photos with or without cues. For example, in one task, which targeted the use of the present continuous to describe actions happening at the time of speaking, the participants were shown a photo of a group of students in a university room doing different activities, and the participants were asked to describe what they saw in that photo. In other tasks, the participants were asked to discuss a topic using a set of vocabulary or phrases. For instance, in one task, which targeted the use of zero article before nouns that refer to classes of people, the participants were shown a list of professions on a card and were asked to classify the professions from the most important to the least important in their views and explain reasons for choosing their classification.

If a brief analysis is done based on R. Ellis' (2005) operationalisation of the constructs of L2 implicit and explicit knowledge, it can be noted that Absi's (2014) speaking test required students (a) to focus on meaning because they had to spontaneously communicate their ideas in a conversation with the researcher; (b) there was a time constraint for each task; (c) the degree of awareness that was required was low because verbalization of their explicit knowledge about the selected grammar points was not required. The results indicate a mean facility value of 70.63% for the speaking accuracy of the participants. With regard to learners' explicit knowledge, the study employed a MLK test to test the participants' metalinguistic knowledge; the test asked participants to identify, correct and explain grammatical errors, and provide sentences exemplifying or illustrating written pedagogical grammar rules. The results indicate that participants found this test somewhat difficult (mean % = 56.13). Analyzing each of the components of the MLK test separately, learners' performance on the rule illustration section was slightly higher (mean % = 67.85) than the error correction (mean % = 66.75). The lowest score was for the rule explanation subsection (25.7%).

In an empirical validation study with 95 participants, Erlam (2006) assesses the oral elicited imitation test that was designed for Ellis' (2006) study (see section 2.1.4). The results indicate that participants corrected 35 percent of ungrammatical statements and a strong positive correlation ($r = 0.73$, $p < 0.00$) between participants' performance at repeating grammatically correct items and their correction of ungrammatical items. In the same study (Erlam, 2006), an oral narrative test was administered. The test presents a short story to read which contains seven target grammar points, a subset of the 17 grammar points included in the elicited imitation task. Participants were asked to read a story twice. The story was removed and they were asked to retell the story in as much detail as possible within three minutes. Correlations were run between the scores of the two tests. There was a significant correlation ($r = 0.48$) between the L2 learners' overall scores on the elicited imitation task and the oral narrative task. Based on the strong positive correlation between participants' performance at repeating grammatically correct items and their correction of ungrammatical items, and the significant correlation between the oral elicited imitation test and the oral narrative test, Erlam (2006) argues that the oral elicited imitation test is reconstructive and a valid measure of implicit knowledge.

In a recent validation study (Kim, Jung, & Tracy-Ventura, 2016), 66 Korean L2 participants completed a Korean elicited imitation test including 30 grammatical sentences, each ranging from 7-19 characters or syllables. A Korean native speaker was digitally recorded speaking the test sentences at a normal rate. A 2-second pause was inserted after each stimulus and before the cue—a 0.5 second beep—which signalled the start of the repetition. Response time length was calculated based on the length of the stimulus (as spoken by the native speaker) plus extra time depending on the number of syllables. Participants also completed the listening section of an international standardized Korean proficiency test including 30 multiple choice items that test skills such as finding a main idea, inferencing, and searching for details. They also completed an adapted independent speaking test item from the Test of English as a Foreign Language (TOEFL). Similar to the TOEFL, participants were given 15 seconds planning time and spoke for approximately 1 minute. The results indicate that there were significant relationships between participants' elicited imitation test scores and the scores of the other two measures. Based on these results, it can be argued that the elicited imitation test is a valid measure of implicit knowledge.

As indicated in the descriptions of the tests above, participants performed differently on tests of implicit and explicit knowledge. In these studies, the imitation test, the oral narrative test, and the timed GJT were designed to measure implicit knowledge because the test takers would rely mainly on feel, they would be under pressure to perform in real time, their focus would be primarily on meaning (with the exception of the timed GJT), and they would not need to access their metalanguage. In contrast, the untimed GJT and the MLK test were designed to measure explicit knowledge because test takers would rely mainly on rule, they would not be under pressure, their focus would be primarily on form, and they would need to access their metalanguage. Table 2.3 summarizes these four criteria.

Table 2.3 Tests of implicit and explicit knowledge

Criterion	**Imitation**	**Oral narrative**	**Timed GJT**	**Untimed GJT**	**MLK test**
Degree of awareness	Feel	Feel	Feel	Rule	Rule
Time available	Pressured	Pressured	Pressured	Unpressured	Unpressured
Focus of attention	Meaning	Meaning	Form	Form	Form
Use of metalanguage	No	No	No	No	Yes

(Adapted from R. Ellis, 2005, p. 157)

A central criticism to explicit and implicit measures is that they might not solely measure what they have been designed to measure, that is, they might lack exclusivity (Ellis, 2005; Gutiérrez, 2013; Rebuschat, 2013: Rodríguez Silva & Roehr-Brackin, 2016; Suzuki & DeKeyser, 2015). In other words, tests of explicit knowledge (e.g. MLK test) may be contaminated by implicit knowledge (Ellis, 2004, 2005; Reingold & Merikle, 1988) if learners are only asked to detect the error in a sentence either in written or spoken form, and conversely, tests of implicit knowledge (e.g. EI test) may be contaminated by explicit knowledge particularly if the stimulus sentence draws learners' attention to form (Erlam, 2006; Susuki & DeKeyser, 2015), especially if participants are explicitly told that the task involves ungrammatical statements (for an overview see Spada, Li-Ju, & Tomita, 2015). One way to avoid such a contamination of implicit knowledge when measuring explicit knowledge is by making learners produce pedagogical grammar rules of the ungrammatical sentences on the test (Gutiérrez, 2013). Likewise, an example may be on the elicited imitation test when learners are asked to repeat an ungrammatical statement correctly pointing to the possibility that learners additionally focus on form, analyze the ungrammaticality of the statement and recur to their explicit knowledge to correct the statement in their minds. One possible way to avoid this second situation is by implementing a word monitoring task (Susuki & DeKeyser, 2015). This task is a computerized task that includes a target word, to which participants need to respond by pressing a button as soon as they hear it. The rationale of the task is that participants slow down to respond to a target word that appears after a grammatical error, which reflects sensitivity to errors. For instance, the response time to the monitored word (i.e., than) in an ungrammatical sentence like "Vicente Fernandez is *more rich than Emmanuel" (where an asterisk designates an ungrammatical element) will be delayed when the monitoring word appears after the ungrammatical part of the sentence, compared to the grammatically correct element (i.e., richer).

Another criticism refers to the criterion of availability of time to perform a task. It is not clear how time pressure can indicate whether a test is more biased to implicit or explicit knowledge because through practice explicit knowledge can become automated explicit knowledge (Bialystok, 1978; Sharwood Smith, 1981) which is functionally equivalent to implicit knowledge (DeKeyser, 2003, 2007; Susuki & DeKeyser, 2015). What this means is that learners can access their explicit knowledge quickly to formulate messages in the L2 and be able to monitor their output when performing spontaneous production tasks (Ellis, 2005, 2006). However, this type of action involving the manipulation and control of information "online" is extremely difficult because working memory can only maintain a very small amount of information (Hulstijn, 2002) for a very short period of time[2].

Despite Hulstijn and De Graaff's (1994) claim that "implicit knowledge is a theoretical construct, not directly accessible by means of language tests" (p. 106), and as discussed in the first criticism above about tests of explicit knowledge (e.g. MLK test) being contaminated by implicit knowledge (Ellis, 2004, 2005; Reingold & Merikle, 1988), R. Ellis' (2005) concrete proposals of how to operationalise implicit and explicit knowledge by means of various tests, as admitted by Hulstijn (2005) a few years later after his claim in 1994, "signals a crucial moment in rendering theories of implicit and explicit knowledge and learning testable" (p. 137). He

[2] The time it takes for information to decay from phonological short-term memory without rehearsal or refreshing of information occurs in the span of 1.5 – 2.0 seconds (Baddeley, Thomson, & Buchanan, 1975).

further states that discussions of construct-definitions claims such as "implicit knowledge of a L2 is what task X measures" might be a topic in SLA research in the near future.

2.1.5 The interface between implicit and explicit knowledge

It is not only the question on how to differentiate one type of knowledge from the other, but also the possible interface between them, which led researchers to adopt one of the following three interface positions: the non-interface position (Hulstijn, 2002; Krashen, 1982; Paradis, 1994), the strong interface position (Anderson, 1982; Bialystok, 1978; DeKeyser, 1998; Sharwood-Smith, 1981), or the weak interface position (R. Ellis, 1993; N. Ellis, 2011).

2.1.5.1 Non-interface position

One of the strongest advocates of the non-interface position is Krashen (1981, 1982, 1985), who differentiates between the concepts of learning and acquisition. He equates learned knowledge with explicit knowledge and acquired knowledge with implicit knowledge and rejects that explicit knowledge can be converted into implicit knowledge. He adopts a non-interface position between the two types of knowledge arguing that (a) L2 acquisition involves subconscious mental processes; (b) learning and acquisition are two different entities; and (c) learners can use their conscious knowledge only to monitor their output.

2.1.5.2 Strong interface position

The strong interface position offers a directly opposing view of the possible interaction between implicit and explicit knowledge. Declarative knowledge can be converted into procedural knowledge by processes of compilation, tuning and restructuring (Anderson, 1982). DeKeyser (1998) refers to tuning and restructuring as automatisation, and he further argues that restructuring can affect both implicit and explicit knowledge. With regard to L2 learning DeKeyser (1998, 2003, 2007) claims that explicit knowledge can be converted into implicit knowledge through communicative drills and intensive practice. DeKeyser argues that only when the learner has the relevant declarative knowledge can s/he engage in practicing a specific grammar point through communicative drills.

From this perspective, it appears that SLA takes place when students acquire declarative knowledge first, that is, when they learn grammar points explicitly. Then, the declarative knowledge can be converted into procedural knowledge through practice. Once the students can make use of the declarative knowledge (i.e. when they have fully understood the use and function of a number of grammar points) they need much meaningful practice to integrate this new knowledge into long-term memory.

In contrast to these two interface positions, the weak interface position takes a compromise view.

2.1.5.3 Weak interface position

According to N. Ellis (2011) explicit learning plays an important role in the development of implicit knowledge in the following ways:

> (a) in the perception of, and selective attending to, L2 form by facilitating the processes of "noticing" (i.e. paying attention to specific linguistic features of the input); (b) by "noticing the gap" (i.e., comparing the noticed features with those the learner typically produces in output); and (c) in output, with explicit knowledge coaching practice, particularly in initial stages, with this controlled use of declarative knowledge guiding the proceduralisation and eventual automatisation of language processing, as it does in the acquisition of other cognitive skills. (p. 36)

These three actions point out that explicit knowledge can indirectly influence how learners develop their implicit knowledge (see N. Ellis, 2011, for an overview; R. Ellis, 2004). As an example of (a) above, while learners may not need help to auditorily notice some salient grammar points such as the progressive *–ing*, the non-salient grammar points such as the third person singular *–s* requires a great deal of effort on the part of the students to notice it. By noticing the non-salient grammar points and repeatedly using the structures in communication is one way that explicit knowledge can indirectly influence how learners develop their implicit knowledge.

It can be noticed that in all interface positions explicit knowledge is a separate entity from implicit knowledge, and these two entities interact differently in each interface position, that is, in the non-interface position explicit knowledge cannot be converted into implicit knowledge through practice. In contrast, in the strong interface position explicit knowledge can be converted into implicit knowledge through intensive practice. In the weak interface position, explicit knowledge through grammar instruction helps learners to notice the L2 grammar points, and by incorporating the grammar points in communicative language use, implicit knowledge may develop. In addition, implicit knowledge may as well help develop explicit knowledge (R. Ellis, 2004, 2005; R. Ellis et al., 2009; Han & Ellis, 1998; Zhang, 2015). This two-way effect between implicit and explicit knowledge may be one central reason why the weak interface position has become the mainstream view nowadays (Dörnyei, 2009). The interest in this topic on the part of the SLA researchers is reflected in the findings of studies on implicit and explicit knowledge.

2.1.6 Findings of studies on implicit and explicit knowledge

Understanding the relationship between implicit and explicit knowledge is of interest to SLA researchers who conduct research to find out its relevance in L2 learning and the role explicit L2 instruction plays in the development of these two types of knowledge (N. Ellis, 2011). The latter issue raises the question as to what extent explicit instruction can improve learners' language proficiency.

2.1.6.1 Correlational studies of implicit and explicit knowledge

Whether explicit knowledge can help with the acquisition of L2 proficiency has been the main focus of several studies (Alderson et al., 1997; Roehr, 2008; Alipour, 2014). Alderson, Clapham, and Steel (1997) report the levels of knowledge about language of first-year undergraduate learners of French and the relationship between this metalinguistic knowledge and language proficiency. Tests of metalinguistic knowledge, and French linguistic proficiency were administered to 509 learners. The results showed a moderate relationship (r = .47) between metalinguistic knowledge and L2 proficiency. The results also showed that learners who scored the highest in the metalinguistic tests did not show the best performance in the target language. Based on these results, the researchers concluded that there was no reason to teach metalinguistic knowledge to improve learners' language proficiency.

More recent studies investigating relationships between L2 proficiency and metalinguistic knowledge have found different results. For instance, in Roehr's (2008) study, tests of metalinguistic knowledge and German linguistic proficiency were administered to first-year and fourth-year university learners (N = 60) of L2 German. The L2 proficiency and metalinguistic knowledge correlated strongly (r = .81) and at a high level of significance for the entire sample of learners. Separate correlations were also calculated for the first-year and fourth-year learners. Surprisingly, the two measures correlated strongly in the advanced learners (fourth-year) and somewhat less strong in the lower-level learners (first-year). Roehr (2008) argues that such a difference may be possibly due to two reasons: (1) "knowledge of grammar and vocabulary as evident in proficient L2 performance may not only be built up on the basis of explicitly acquired metalinguistic knowledge, but may also help a learner develop their metalinguistic knowledge" (p. 192) and (2) due to the combined impact of factors such as cognitive ability, motivation, and attitudes toward formal language study.

In a more recent study (Alipour, 2014) a test of metalinguistic knowledge and a test of L2 proficiency were administered to 38 English as a foreign language (EFL) learners at the university level. Twenty learners were in their first year of study and 18 were in the second year of their study. The two tests correlated moderately (*r* = .53). The researcher argues that a plausible explanation for such a relationship is due to the inclusion of the same type of items in both tests, which may help explain why learners have implicit and explicit knowledge of specific grammar points.

In another study (Absi, 2014 – see section 2.1.4), the researcher calculated correlations between participants' explicit knowledge as measured by a MLK test and implicit knowledge as measured by a speaking test. The participants' scores on the speaking test as a whole moderately correlated (*r* = .397) with their scores on the MLK test. In terms of perceived difficulty (i.e. easy, moderate, and difficult grammar points) the relationship between implicit and explicit knowledge of the grammar points as a whole ranged from weak (*r* = .267) to moderate (*r* = .402). The strongest correlation was at the level of difficult grammar points. Absi (2014) argues that this relationship is "an indirect piece of evidence on the effectiveness of explicit FonFs grammar instruction" (p. 225).

An analysis by grammar point yielded different results in two studies (R. Ellis, 2006; Rodríguez Silva & Roehr-Brackin, 2016). In R. Ellis' (2006) study (see section 2.1.4) the researcher ran a Spearman rank order correlation between implicit and explicit scores for the 17 individual grammar points, and he found a very weak and statistically non-significant correlation (*r* = 0.08). He also ran a correlation between the implicit/explicit knowledge scores for the seventeen grammatical structures and the IELTS (International English Language Testing System) scores (total, listening, speaking, reading and writing) and significant correlations were found. R. Ellis (2006) argues that "implicit knowledge of the grammatical features was more strongly related to oral IELTS while the reverse was the case for explicit knowledge" (p. 458). Thus, what these correlations show is that structures that are easy in terms of implicit knowledge may be difficult in terms of explicit knowledge and vice versa.

In Rodríguez Silva and Roehr-Brackin's (2016) study, 30 intermediate-level learners at a higher-education institution in Mexico completed tests of explicit and implicit L2 knowledge. The test of implicit L2 knowledge was an elicited imitation test (R. Ellis, 2006; R. Ellis et al., 2009; Erlam, 2006), and the test of explicit L2 knowledge was a metalinguistic knowledge test (Absi, 2014; Scheffler, 2011). A correlational analysis showed that scores on the MLK test and the EI test were associated (r = 0.65), suggesting that instructed L2 learners develop both implicit and explicit knowledge to at least some extent. A further analysis by targeted grammar point found no significant association between implicit and explicit knowledge suggesting that for a particular L2 construction learners develop one type of knowledge first and subsequently construct the other type of knowledge, rather than acquiring both types of knowledge together. In terms of the learners' performance on the measures of implicit and explicit knowledge, one finding that is shared between Ellis' (2006) study and Rodríguez Silva and Roehr-Brackin's (2016) study is that the grammar points that were discovered to be easy for learners in terms of their explicit knowledge were difficult in relation to their implicit knowledge and vice versa.

Further evidence to support what type of role explicit instruction plays in the development of implicit and explicit knowledge can be found in experimental studies in which measures of implicit and explicit knowledge are used.

2.1.6.2 Experimental studies on implicit and explicit knowledge

In an early experimental study (Fotos, 1993), 160 Japanese university EFL learners were randomly assigned to three treatment groups consisting of 53 to 54 learners. The three treatment groups were as follows: (1) a grammar task group, who performed three grammar consciousness-raising tasks ("consciousness-raising" refers to increased learner awareness of particular linguistic features (Rutherford & Sharwood Smith, 1985)); (2) a grammar lesson group, who received grammar lessons identical in content to the grammar tasks, and (3) a communicative task group, who performed communicative tasks matched in format, length, and task features, but lacking grammatical content. Before the research began, the three classes were administered a cloze test; the results of this test indicated that the three classes were considered to be equivalent. The learners had one required 90-minute period per week of oral English with a native speaker instructor. The tasks/lessons were administered in three cycles of three weeks each. During the first week of each cycle, the two task groups performed the tasks and the grammar lesson group received a formal, teacher-fronted lesson, the contents of which were taken directly from the task sheets and task cards used for the grammar task. Before performance of the grammar tasks and administration of the grammar lessons, both grammar treatment groups took pre-tests on the targeted grammar structures (indirect object placement, adverb placement and relative clause usage). After the tasks/lessons, the two grammar treatment groups took post-tests which were identical to the pre-tests; the learners in the communicative task group, which served as the control group, did not take any of the tests so as not to expose them to the target structures. The second week after the task performance/grammar lesson, the two treatment groups and the control group were read a story of about 150 words, with the target structures from the previous week's treatment embedded five times within the story. After listening to the story and answering several general questions about the contents, the groups were given the written texts and were asked to check their answers and read the story, underlining anything which they considered special or noteworthy of some grammatical structures they noticed. After five minutes the texts were collected. During the third week, all three groups were given a dictation exercise with the target structures embedded three to four times in 75-80 words. After the groups had written the dictation, they were given the texts and asked to read the dictation and check what they had written. They were asked to underline any grammatical structures they noticed, and the texts were collected after five minutes. At no time during the noticing exercises did the researcher comment on the presence of the target grammar structures. Frequencies were obtained by counting the number of times the structure was underlined in the noticing exercises. The frequencies of noticing the target structure in communicative input one and two weeks after the grammar-consciousness treatments were compared with the noticing frequencies of the control group which was not exposed to any type of grammar consciousness-raising activity.

When a correlational analysis was run for both treatment groups and the control group, only the grammar-lesson group showed a low but significant positive correlation between the noticing frequencies and the final proficiency test scores for the grammar point indirect object. Put differently, the significant correlation points to the relationship between the frequency of noticing a grammar point and proficiency gains in the use of that grammar point. In other words, there seems to be an interface between explicit knowledge and implicit knowledge of the same grammar point. The author (Fotos, 1993) argues that a plausible explanation for this type of result seems to be "some sort of threshold effect operating for noticing levels, whereby a certain frequency of noticings must occur before consolidation of explicit knowledge takes place" (p. 400) and this may possibly be an indication of the development of implicit knowledge of target grammar points.

In a more recent study (Akakura, 2012), an elicited imitation task, an oral production task, and a MLK test were administered to 98 participants in an experimental study. Participants were randomly assigned to the experimental group (n = 49) or control group (n = 45). Participants in the experimental group were exposed to instruction delivered through computer-assisted language learning (CALL) activities on a website that can be used by learners on their own. The control group did not receive instructional treatment. There were three testing times: a pre-test, an immediate post-test, and a delayed post-test six weeks after treatment. The target grammatical forms were generic (indefinite *a/an*, and the definite *the*) and the non-generic use of the English article (indefinite *a*, *an,* and definite *the*). The EI test consisted of 14 true or false sentences in which grammatical (*n* = 10) and ungrammatical (*n* = 10) articles were tested. Participants were asked to listen to a story while looking at a series of 20 pictures illustrating it. Based on these, the researcher created narratives embedding target articles. Whereas the pictures depicted the true story, half of the recorded narrative contained sentences that were contrary to the pictures; at the end of each sentence participants heard the question "Is this true or false?" Participants were given an answer sheet where they could tick their answers: true, false or not sure. The "True or False" was intended to focus attention on meaning. Participants then were asked to describe the picture orally using all the words provided under the picture. An example is provided below (Akakura, 2012).

Recording: "*During the night when no one was looking, unexpected thing happened. Is this true or false?"

Picture of a frog stepping out of a glass jar at night, while the boy and dog are fast asleep on the bed.

Words provided: During the night when…, unexpected thing…(p. 18)

On the oral production task, participants had to narrate a story for the same sequence of pictures as the elicited imitation task, but in their own words; no words were provided for this task. On the MLK test participants were asked to correct 10 sentences each containing an article error that was underlined (*n* = 10). Then they had to explain the correction; there were no time constraints to complete the test. Results of a one-way ANOVA revealed that the experimental group outperformed the control group on all components of the post-tests with mostly large effect sizes. The findings show that explicit instruction has an effect on implicit knowledge. This means that instructed learners can develop implicit knowledge of non-salient grammar points. The findings also show that explicit instruction has an effect on explicit knowledge concerning ungrammatical stimuli only. This fact suggests that teachers should carefully plan activities to develop learners' explicit and implicit knowledge. Hence, these results show evidence that explicit instruction can lead to the development of both explicit and implicit knowledge.

On the same topic, three meta-analyses reveal interesting results regarding the role of explicit instruction in L2 learning with regard to the kind of knowledge this type of instruction generates (Spada, 2011), that is, explicit and/or implicit knowledge. Norris and Ortega's (2000) meta-analysis involving 49 experimental and quasi-experimental studies published between 1980 and 1998 indicates that focused L2 instruction results in large target-oriented gains, specifically, explicit types of instruction are more effective than implicit types, and the effectiveness of L2 instruction is durable. These results were replicated in Goo, Granena, Yilmaz, and Novella (2015) study involving 34 unique sample studies: 11 studies from Norris and Ortega's (2000) meta-analysis and 23 new studies published between 1999 and 2011. In another meta-analysis, Spada and Tomita (2010) investigated the effects of explicit and implicit instruction on the acquisition of simple and complex grammatical features in English. The results indicate larger effect sizes for explicit over implicit instruction for simple and complex features. The findings also suggest that explicit instruction positively contributes to learners' controlled knowledge and spontaneous use of complex and simple forms.

Related to the results in the three meta-analyses, Cerezo, Caras, and Leow (2016) argue that these meta-analyses subsumed both deductive and inductive instruction as explicit instruction, but these meta-analyses "did not clarify which type of explicit instruction yields the best results" (p. 266). In their study (Cerezo, Caras, & Leow, 2016), which targeted the complex Spanish *gustar* structures, 70 English-speaking learners of beginning Spanish received either guided induction (n = 24) via a videogame, deductive instruction (n = 26) in a traditional classroom setting, or no instruction (n = 20). In the guided induction group, on starting the game, participants were greeted by a pop-up bubble asking, "*¿Cómo se dice en español* (How do you say in Spanish) I like the house*?*" The screen then displayed two options, *yo* (I) and *a mí* (to me). Participants selected one, received corrective feedback, and were presented with the next two options to complete the sentence. The treatment consisted of 20 items like this one, sequentially presented according to four types of *gustar* structures, across four video game levels. After successfully completing a video game level, the screen displayed a list of correct and incorrect rules for the *gustar* item types. Participants were asked to select those that they thought were correct and were taken to the next level regardless of their accuracy. In the deductive instruction group, the teacher covered all 20 exemplars of *gustar* described in the videogame in the same order and explained the rules. The control group performed the assessment tasks without any formal exposure to the targeted structure. Assessment tasks included two productive tasks (controlled oral and written production), which measured participants' ability to produce orally and in writing the targeted *gustar* structure, and one receptive task (multiple-choice written recognition). The latter tests were administered 2 weeks after the immediate posttests. With respect to the oral production and controlled production task, the results indicate that both experimental groups experienced statistically significant learning from pretest to immediate posttest, and the guided group retained learning from the immediate to the delayed posttest while the deductive instruction group experienced a statistically significant decrease. The control group did not improve significantly from pre-test to post-test, but it improved significantly from the immediate to the delayed post-test. The results also indicate that the experimental groups outperformed the control group, the guided instruction group was the top scorer, but it did not significantly outperform the deductive group, and at the delayed post-test, only the guided instruction group outperformed the control group.

With regard to the multiple choice written recognition task, both experimental groups improved significantly from the pre-test to the immediate post-test, and they retained learning from the immediate to the delayed post-test as well. These results loosely demonstrate that explicit instruction (whether guided induction or deductive instruction) has an effect on explicit L2 learning, and that deductive instruction effectively promotes the development of explicit knowledge of difficult grammar points.

The findings presented in this section on the role that explicit instruction plays in the development of L2 learning concerning explicit and implicit knowledge demonstrate that some structures are "harder to notice without explicit focus on form" (DeKeyser, 1998, p. 46), and that these empirical studies (and others) also indicate that explicit instruction leads to higher L2 learning gains mainly on explicit knowledge and to some extent on implicit knowledge. In the same vein, both research and teaching experience are leading to a consensus that explicit instruction, either inductive instruction or deductive instruction is most effective when it includes attention to form and meaning (Spada & Lightbown, 2008) which is reflected in learners' development of explicit and implicit knowledge.

2.1.7 Interim summary

In this section about implicit and explicit knowledge and learning in SLA, key characteristics and the operationalisation of the constructs of implicit and explicit knowledge have been proposed by R. Ellis (2005). So far, no other theoretical perspective or empirical finding has proposed something different. This fact, along with the validation studies (R. Ellis, 2006; Erlam, 2006; Gutiérrez, 2013) and empirical studies (Absi, 2014; R. Ellis, 2005; Rodríguez Silva & Roehr-Brackin, 2016; Roehr & Gánem-Gutiérrez, 2009b) supports R. Ellis' (2005) proposals.

It was also discussed that the use of implicit measures (e.g. the elicited imitation test and the timed grammaticality judgement test) has increased in empirical studies in recent years (Goo et al., 2015; Spada & Tomita, 2010), and that the results of explicit measures (e.g. the untimed grammaticality judgement test and the MLK test) showed a significant difference in comparison to implicit measures. This raises the issue on the learning mechanisms learners develop depending on the type of instruction they are instructed with. That is, it is probable the type of instruction was deductive instruction and consequently, this contributed to learners' explicit performance on the explicit measures.

The issue in relation to the efficiency of timed and untimed grammaticality judgement tests as measures of explicit knowledge is not clear given the different findings in a number of studies (R. Ellis, 2006; Gutiérrez, 2013). It appears that there is more conclusive evidence for the MLK test as a measure of explicit knowledge (Rodríguez Silva & Roehr-Brackin, 2016; Ziętek & Roehr, 2011; Scheffler, 2011). No issues have been raised with regard to the EI test or speaking tests as measures of implicit knowledge.

Another issue raised in this section was to what extent explicit instruction can improve learners' language proficiency, and the empirical findings suggest that the teaching of metalinguistic knowledge does improve learners' language proficiency (Fotos, 1993; Akakura, 2012; Alipour, 2014; Absi, 2014; Cerezo et al., 2016; Roehr, 2008), though not all studies share similar findings (see Alderson, apham, & Steel, 1997). These results demonstrate that explicit knowledge correlates with implicit knowledge to some extent, and the results of analysis by grammar point show that both types of knowledge do not occur at the same time, that is, learners acquire one type of knowledge first and then the other. This is an important finding because it suggests that teachers should think carefully whether to follow the class textbook page by page or plan their activities according to what structures learners may struggle more and spend more time on difficult structures.

Needless to say, for the past three decades, SLA research has not only focused on the relationship between implicit and explicit knowledge but also on the learning difficulty of grammar points (Collins et al., 2009; DeKeyser, 2005; R. Ellis, 2006; Dietz, 2002;

Goldschneider & DeKeyser, 2001; Graus & Coppen, 2015; Housen, Pierrard, & Van Daele, 2005; Hulstijn & De Graaff, 1994; Palloti, 2015; Rodríguez Silva & Roehr-Brackin, 2016; Scheffler, 2009, 2011; Spada & Tomita, 2010). This latter issue is discussed in the section that follows.

2.2 Learning difficulty in instructed L2 learning

Learning difficulty in SLA related to grammar points is defined as "the mental ease or difficulty with which linguistic items are learned, processed or verbalized in the processes of language acquisition and use" (Bulté & Housen, 2012, p. 23). Two categorizations of difficulty are worth mentioning because they are most relevant in the present context, that is, the levels of difficulty of cognitive processes among individuals and the levels of difficulty of linguistic grammar points (DeKeyser, 2003; Housen & Simoens, 2016). Put differently, the concept of learning difficulty can be approached from a subjective perspective (processing difficulty or cognitive difficulty) and/or an objective perspective (structural complexity of the linguistic constructions). In the latter categorization, the notion of difficulty is often used as a synonym for structural complexity (Dietz, 2002).

2.2.1 Structural complexity as number of transformations

The concept of structural complexity has been discussed from different approaches: psycholinguistics, linguistics, and pedagogy (Spada & Tomita, 2010). With respect to the psycholinguistic approach, Spada and Tomita (2010) argue that a grammar point that is acquired early must be easy to learn, whereas a grammar point that is acquired late must be difficult to learn and this in itself is circular in nature. The linguistic approach is also problematic because what is easy to describe is not always easy to learn (e.g., third-person singular –*s*), and a pedagogical approach poses the situation that different learners may find some grammar points more difficult than others.

Another approach used to determine the difficulty of grammar points can be found in Spada and Tomita's (2010, p. 269) meta-analysis. This meta-analysis comprises thirty experimental studies on the effects of instruction on the acquisition of simple and complex 12 English grammar points. The studies included were laboratory studies and focused on complex grammar points, and studies conducted in classroom contexts on both complex and simple grammar points. The main purpose of the study was to examine the effects of different types of instruction on L2 learning. To categorize the 12 grammar points into complex or simple the researchers employed Hulstijn and de Graaff's (1994) definition of structure complexity according to which "the degree of complexity is contingent on the number (and /or the type) of criteria to be applied in order to arrive at the correct form" (p. 103). Table 2.4 provides an example of two grammar points classified as complex and simple by applying this criterion.

Table 2.4 Number of transformations: Complex and simple rules

Complex rule: ***Wh*-question of an object of preposition** **Example: "Who did you talk to?"**	**Simple rule:** **Regular past tense** **Example: "walked"**
1. *Wh*-replacement *(You [past] talk to who)*	1. [Past tense] + Verb
2. *Wh*-fronting *(Who you [past]talk to)*	[Past tense] + walk = walked
3. Do support *(Who you [past]do talk to)*	
4. Subject/auxiliary inversion *(Who [past] do you talk to)*	
5. Affix attachment *(Who [DO + past] you talk to)*	
6. Morphological rules *(Who did you talk to)*	
7. Fronting/leaving behind *(To whom did you talk?/Who did you talk to?)*	

(Spada & Tomita, 2010, p. 272)

Table 2.4 shows how the grammar point "*Wh*-question of an object of preposition" is classified as "complex" because it requires more than one transformation to arrive at the sample sentence *Who did you talk to?* than the past tense of the regular verb *walk* which is classified as "simple" because it only requires one transformation to arrive at the target form: suppliance of the *–ed* inflection.

The coding for instruction as explicit was based on the criteria (a) when it included grammar rule explanation, (b) L1/L2 contrasts, and (c) metalinguistic feedback. The coding for instruction as implicit was based on (a) input flood/high-frequency, (b) interaction, and (c) recasts. The results indicate that the effects of explicit instruction are statistically significant for complex and simple features. In contrast, the effects of implicit instruction are significant for complex features only. The researchers stated that if they had chosen different criteria to differentiate between simple and complex L2 features, they might have obtained different results taking into account that classifying a number of grammar points on the basis of a single learning difficulty criterion may leave out factors that have a direct impact on implicit and explicit learning such as frequency of input (Larsen-Freeman & Long, 1991; R. Ellis, 2006) and technicality of the grammar point in question (R. Ellis, 2006; Dietz, 2002). Another problem with this type of operationalisation of structural complexity is that a number of structures classified as easy (i.e., structures that require one or two transformations according to Spada and Tomita's (2010) classification) are difficult to learn, such as possessive determiners, indefinite articles, and past tense of regular verbs as evidenced in SLA (Collins et al., 2009; Ellis, 2006; Erlam, 2006). One more problem is that this approach makes no distinction between implicit and explicit knowledge in the sense that it does not differentiate between linguistic constructions and metalinguistic descriptions. Furthermore, classifying grammar points with this type of operationalisation leaves out subjective factors (DeKeyser, 2003; R. Ellis, 2008; Dörnyei, 2005; Gilabert & Muñoz, 2010; Graus & Coppen, 2015) such as developmental stage, language aptitude, and working memory capacity of individuals as well as other variables such as perceptual salience, and communicative redundancy that pertain to implicit and explicit learning difficulty.

2.2.2 The cognitive difficulty perspective

It would be advantageous if an approach could be singled out and used to accurately predict the learning difficulty of different L2 grammar points and their associated pedagogical grammar rules. Despite the efforts by a number of researchers to find out what makes the acquisition of different grammar points more or less difficult, it is an area in which no consensus has been yet reached in the research literature (Collins et al., 2009; DeKeyser, 2005; Housen & Simoens, 2016; Hulstijn & de Graaff, 1994; R. Ellis, 2006; Scheffler, 2011; Robinson, 1996; Spada & Tomita, 2010). Krashen (1982) and R. Ellis (1990), for instance, disagree on the learning difficulty of the grammar point third person –*s*. The former classifies the grammar point as formally simple referring to the presence or absence of a single morpheme, whereas the latter classifies it as formally complex pointing to the form-function relationship between the grammatical number of the subject and the presence or absence of the morpheme –*s* on the verb. The lack of agreement on the difficulty of the third person –*s* between these two researchers is probably due to the narrow or broader sense (DeKeyser, 2005) each researcher classifies the grammar point, that is, Krashen (1982) may have judged the difficulty of this grammar point based on whether the form of the verb includes or precludes the suffix –*s*, whereas R. Ellis (1990) may have focused on the distance between the co-occurring elements such as noun phrase complements or adverbs between the subject and the verb, and thus make this subject-verb relationship difficult to acquire (DeKeyser, 1998). In contrast, DeKeyser (1998) regards this grammar point to be functionally complex given the number of grammatical categories involved in its interpretation such as person, number, and tense. DeKeyser (2005) proposes that grammatical difficulty of grammatical structures may be due not only to complexity of form as suggested by Krashen (1982), or the form-function relationship suggested by R. Ellis (1990), but rather to the lack of transparency of form-meaning relationships that determines the difficulty of grammar points, and by extension, the difficulty in their acquisition.

2.2.2.1 The relative difficulty of grammar points as implicit knowledge

In the preceding approaches, the topic of linguistic or structural complexity was dealt in terms of a single criterion (i.e. number of transformations) (Spada & Tomita, 2010), form-function relationships (Ellis, 1990), and lack of transparency of form-meaning relationships (DeKeyser, 2005). In this section, this topic is addressed in terms of a specific perspective trying to depict how different researchers view the learning difficulty of grammar points as implicit knowledge. R. Ellis (2006: 435) drawing on the work of N. Ellis (1996), Goldschneider and DeKeyser (2000), Hulstijn and de Graaff (1994) and Pienemann (1999) proposed the following criteria of structural difficulty as implicit knowledge of different grammar points:

1. Frequency (i.e. How frequently does the grammatical feature occur in the input?)
2. Perceptual salience (i.e. Is the grammatical feature easy to notice in the input?)
3. Functional value (i.e. Does the grammatical feature map onto a clear, distinct function?)
4. Regularity (i.e. Does the grammatical feature conform to some identifiable pattern?)
5. Processability (i.e. Is the grammatical feature easy to process?)

Notwithstanding that these criteria can be useful to determine structural difficulty as implicit knowledge of different grammar points, R. Ellis himself observes that it is not clear how to apply such criteria to determine the relative learning difficulty of different grammar points. As far as the applicability of these criteria is concerned, Roehr and Gánem-Gutiérrez' (2009a) taxonomy (henceforth: R&GG taxonomy, see Table 2.5) provides an example on how to apply some of the variables of R. Ellis' criteria in empirical research (Huang, 2012; Rodríguez Silva & Roehr-Brackin, 2016; Thepseenu & Roehr, 2013; Ziętek & Roehr, 2011).

By the same token, DeKeyser (2005) proposed that three factors are responsible for the lack of transparency in the link between form and meaning:

1. Communicative redundancy (i.e. The form at issue is not semantically necessary because its meaning is also expressed by at least one other element of the sentence)
2. Optionality (i.e. The presence or absence of an element either at *sentence level* such as null subject in Spanish or *word level* such as case marking in Korean — emphasis added)
3. Opacity (i.e. Different forms stand for the same meaning, and the same form stands for different meanings) (p. 8)

Although DeKeyser (2005) does not make any observation on the applicability of these three factors to determine the learning difficulty of different grammar points as implicit knowledge, it is not hard to read between the lines to see that he agrees with R. Ellis (2006) that it is not yet clear how to apply such criteria in empirical research. Moreover, learning difficulty of grammar points is not only about learning them through implicit processes but also through explicit processes because the relative learning difficulty of any grammar point involves the characteristics of both its linguistic construction and metalinguistic description.

2.2.2.2 The relative difficulty of grammar points as explicit knowledge

In an effort to explain the learning difficulty of grammatical structures concerning explicit learning difficulty, R. Ellis (2006) also proposed two criteria of difficulty as explicit knowledge of different grammar points:

1. Conceptual complexity (i.e. This concept is defined as the number of different formal or functional grammatical features that contribute to the specific form of a target structure and the specific functions it performs)
2. Technicality of metalanguage (i.e. This concept can be "semi-technical" or "technical") (pp. 438-39)

Following these two criteria of explicit learning difficulty, Roehr and Gánem-Gutiérrez (2009a) included in their taxonomy (see Table 2.5) two more factors that may influence how easy or difficult a grammar point can be: schematicity and truth value.

2.2.2.3 Taxonomy of implicit and explicit learning difficulty of grammar points

As mentioned above, the applicability of criteria to determine the cognitive learning difficulty of grammar points as implicit and explicit knowledge is exemplified by Roehr and Gánem-Gutiérrez (2009a) who drew on both DeKeyser (2005) and R. Ellis, (2006) to develop a taxonomy for assessing the implicit and explicit learning difficulty of the L2 grammar points included in their study. The researchers used their taxonomy to assign a low, medium, or high value to each targeted L2 construction and its associated metalinguistic description to determine implicit and explicit learning difficulty. The taxonomy comprises the variables frequency, perceptual salience, communicative redundancy, and opacity, and they refer to the characteristics of linguistic constructions and impact on implicit learning difficulty. The variable schematicity refers to the characteristics of both linguistic constructions and metalinguistic descriptions and affects both implicit and explicit learning difficulty. The variables conceptual complexity, technicality of metalanguage, and truth value refer to the characteristics of metalinguistic descriptions and impact on explicit learning difficulty as shown in Table 2.5.

Table 2.5 Taxonomy of variables contributing to implicit and explicit learning difficulty

Variable	Operational definition	Learning difficulty
Frequency	How frequently an L2 construction occurs in the input.	High frequency decreases implicit learning difficulty.
Perceptual salience	How easily an L2 construction can be perceived auditorily in spoken input.	High perceptual salience decreases implicit learning difficulty.
Communicative redundancy	How much an L2 construction contributes to the communicative intent of a message.	High communicative redundancy increases implicit learning difficulty.
Opacity (lack of	To what extent an L2 form maps	High opacity increases implicit

reliability) of form-meaning mapping: One form, x meanings	onto a single or multiple meanings/ functions.	learning difficulty.
Opacity (lack of reliability) of meaning-form mapping: One meaning x forms	To what extent an L2 meaning/ function maps onto a single or multiple forms.	High opacity increases implicit learning difficulty.
Schematicity	The extent to which a linguistic construction is schematic or specific; and whether a meta-linguistic description covers a schematic or a specific linguistic construction.	High schematicity decreases implicit and explicit learning difficulty.
Conceptual complexity	The number of elements that need taken into account in a meta-linguistic description, i.e. the number of categories and relations between categories included in the description.	High conceptual complexity increases explicit learning difficulty.
Technicality of metalanguage	The relative familiarity and abstractness of the metalanguage used in the metalinguistic description.	High technicality of metalanguage increases explicit learning difficulty.
Truth value	The extent to which a meta-linguistic description applies without exception.	High truth value decreases explicit learning difficulty.

(Adopted from Roehr & Gánem-Gutiérrez, 2009a, p. 88)

While the R&GG taxonomy of variables contributing to implicit and explicit learning difficulty appears to be the only taxonomy of its kind, its application may require careful administration. On one hand, judging each variable in the taxonomy may not be straightforward. Namely, the variables perceptual salience, communicative redundancy, and opacity are notions that L2 teachers and researchers can understand fairly easily, but making accurate judgements for each of the criteria in the taxonomy presents the problem that teachers and researchers have to rely on their intuitions and knowledge about the learning difficulty of grammar points which may or may not be reliable.

In a number of studies (Huang, 2012; Rodríguez Silva & Roehr-Brackin, 2016; Thepseenu & Roehr, 2013; Ziętek & Roehr, 2011) where difficulty judgement questionnaires for L2 experts were used, applied linguists were asked to assess learning difficulty as implicit and explicit knowledge in accordance with the R&GG taxonomy and to assign a value of "high", "medium" or "low" to each variable in the taxonomy for each of the 13 targeted grammar points. An example is given below (Rodríguez Silva, 2013).

Example:

1) Simple past tense

When a finished action or event in the past is being expressed, the past tense is required.
He visit**ed** his brother yesterday.
*When he finished his homework he <u>watch</u> a movie.

Variable	Value
1) Frequency	
2) Perceptual salience	
3) Communicative redundancy	
4) Opacity: One form, X meanings	
5) Opacity: One meaning, X forms	

Variable	Value
6) Schematicity	
7) Conceptual complexity	
8) Technicality of metalanguage	
9) Truth value	

These studies were conducted on different populations and contexts and have yielded mixed results as outlined in the following.

In Huang's (2012) study, for instance, three Taiwanese researchers (including the researcher herself) completed a difficulty judgment questionnaire based on the R&GG taxonomy. The researchers evaluated twelve grammar points by assigning one of the three values (high, medium, low) to each criterion in the taxonomy. The inter-rater reliability was below 70%. The researcher participants, as it was found in the study, assigned the three values according to different subjective standards. Another situation was the comments of two researchers on the grammar points of the study; they found that there was ambiguity in some grammar points which rendered hard to decide the form and meaning these grammar points represented, making the evaluation of the two opacity criteria difficult. The main researcher of the study had to rephrase some grammatical rules and substitute some grammar points by more clear-cut grammar points to avoid discrepancy on judgements. The procedure of assessment was also modified by presenting guiding questions first and then asking the participants to assign three values to each criterion for the example grammar points according to the guiding questions. These changes increased the inter-reliability.

In Rodríguez Silva and Roehr-Brackin's (2016) study, three applied linguists (including the first author of the study), 11 teacher participants, and 30 student participants were asked to assess learning difficulty of thirteen grammar points. The applied linguists assessed the learning difficulty of the thirteen grammar points as implicit and explicit knowledge based on the R&GG taxonomy, and the teacher and student participants assessed the same grammar points based on a 5-point scale (1 = very easy, 2 = easy, 3 = moderate, 4 = difficult, 5 = very difficult). The results show no significant correlations between the applied linguists' judgements and learners' judgements, and between the applied linguists' judgements and learners' performance on implicit and explicit measures. The dissertation (Rodríguez Silva, 2013), on which Rodríguez Silva and Roehr-Brackin's (2016) study is based, accounts for the difficulty that two expert participants had to understand how to proceed in their difficulty judgements of each grammar point. An average of three hours was spent with each expert participant because they found the operational definition of some variables hard to understand. A second meeting took place with both of them to start working on the questionnaires to apply the taxonomy and evaluate the learning difficulty of the grammar points. The second meeting took two hours. In that section only half of the grammar points were evaluated, and they agreed on completing the rest of the grammar points individually and hand the questionnaires to the first author of the study later on. Both participants on three different occasions approached the researcher to clarify the operational definition of some particular criteria. The process of administration of this questionnaire suggested that the applied linguists may not have been able to use the criteria reliably to evaluate the targeted grammar points as easy, moderate, or difficult.

In contrast to these two studies, the R&GG taxonomy of learning difficulty was employed successfully to differentiate between lower and higher explicit learning difficulty of various grammar points in two further studies (Thepseenu & Roehr, 2013; Ziętek & Roehr, 2011). In the latter, Ziętek and Roehr (2011) found that Polish college-level learners (N = 20) performed significantly worse on metalinguistic test items targeting six grammar points judged by the two authors of the study to be of higher explicit learning difficulty than on items targeting six grammar points judged to be of lower explicit learning difficulty based on the R&GG taxonomy. Ziętek and Roehr's (2011) finding was replicated in Thepseenu and Roehr's (2013) study. This study used the same test with Thai university-level learners (N = 64). Hence, more studies in this line of research may elucidate whether the use of the taxonomy can be a fruitful approach to determine the cognitive learning difficulty of grammar points as implicit and explicit knowledge.

An alternative approach in determining the ease or difficulty of grammar points is asking L2 teachers to make holistic judgements on the learning difficulty of grammar points based on their experience in teaching them and their intuitions about the learning difficulties learners go through when trying to learn them (Absi, 2014; Graus & Coppen, 2015; Robinson, 1996; Scheffler, 2011).

2.2.3 Holistic judgements of learning difficulty of grammar points

While some researchers (Spada & Tomita, 2010) have attempted to classify a number of grammar points on the basis of a single learning difficulty criterion, and others (Roehr & Gánem-Gutiérrez, 2009a) have designed a taxonomy of variables contributing to implicit and explicit learning difficulty of linguistic constructions and their associated pedagogical grammar rules, a different approach to determine the difficulty of grammar points was employed in various studies (Graus & Coppen, 2015; Robinson, 1996; Rodríguez Silva & Roehr-Brackin, 2016; Scheffler, 2011). These studies used the expert judgements of experienced L2 teachers based on subjective criteria such as their knowledge on how the target language system works, on their understanding of the cognitive processes required to learn the target language, and on their intuitions on how learners struggle to acquire the grammar points of the target language.

In Robinson's (1996) study, a number of pedagogic rules were given to 15 experienced ESL teachers to rate for complexity using rating scale judgments of rule complexity. It is worth noting that the rating scales to judge the learning difficulty of grammar points are very simple in nature, that is, the values range from easy to difficult, but the criteria that teachers were expected to use is not spelled out in the questionnaire itself. The approach teachers in Robinson's (1996) study followed identified two rules and their corresponding structures as differing in complexity: a difficult rule for describing how to form pseudoclefts of location (e.g. *Where Mary and John live is in Chicago not in New York*), and an easy rule describing the fact that subject-verb inversion is allowed in sentences where adverbials of movement or location are fronted (e.g. *Into the house John ran/ran John*). The study addressed Reber's (1989, 1993) and Krashen's (1981, 1982, 1985, 1994) claims that (a) implicit learning is more effective than explicit learning when the stimulus domain is complex, and (b) explicit learning of simple and complex stimulus domains is possible if the underlying rules are made salient. The learners in this study (N = 104) were randomly assigned to one of the four computerized training conditions: implicit, incidental, rule-search, or instructed. The implicit condition was explained to participants as a memory test. Participants viewed sentences conforming to the two rules and were instructed to remember them. The incidental condition was explained as an exercise in reading for meaning, whereas the rule-search condition was explained as an exercise in identifying the rules illustrated by sentences. In the instructed condition participants read through the rules that were the focus of the study. After completing the training session, participants were asked to complete a grammaticality judgement task on the computer. They were instructed to respond as quickly as they could to each sentence by indicating whether the sentence was grammatical or ungrammatical. The results do not support Reber's and Krashen's claim that implicit learning is more effective than explicit learning when the grammar points are complex, but the results do support the second claim that explicit learning of simple and complex stimulus domains is possible if the underlying rules are made salient. In other words, implicit learners did not outperform other learners on difficult structures, but instructed learners did outperform all other learners in learning simple structures.

In another study, Scheffler's (2011) results show a very strong significant correlation (rho = -.9, $p < .01$) between Polish secondary school teachers' ($n = 25$) difficulty judgements of 12 grammar points relating to the form and meaning of the L2 English verb phrase on a five-point scale ranging from "very easy" to "very difficult", and their students' ($n = 50$) performance on a metalinguistic test requiring the production of English sentences exemplifying the targeted pedagogical grammar rules.

Rodríguez Silva and Roehr-Brackin (2016) conducted a study with applied linguists ($n = 3$) (including the first author of the study) with postgraduate-level qualifications, university teachers of English ($n = 11$), and learners ($n = 30$) at a higher-education institution in Mexico as described above. The instruments used in the study were (a) difficulty judgement questionnaires for the three groups of participants and (b) tests of explicit and implicit L2 knowledge for the learners. All instruments presented 13 targeted grammar points following the pedagogical grammar rules in the format: “When form X occurs/function X is being expressed, form Y needs to be used” (Roehr, 2008; Thepseenu & Roehr, 2013; Ziętek & Roehr, 2011). The difficulty judgement questionnaires for teachers and learners consisted of the grammar point, the associated pedagogical grammar rule, an example sentence illustrating the use of the linguistic structure, and an example of a typical learner error. Participants were asked to indicate their opinion about the difficulty of each grammar point on a five-point scale (very easy – easy – moderate – difficult – very difficult) (DeKeyser, 2003), based on their experience of teaching English (teachers) or learning English (learners) as shown in the following example:

<table>
<tr><th rowspan="2">Grammar Point</th><th rowspan="2">Pedagogical grammar rule</th><th rowspan="2">Example sentence(s) (targeted form is in bold)</th><th rowspan="2">Typical learner error (error is underlined)</th><th colspan="5">Level of difficulty</th></tr>
<tr><th>Very easy</th><th>Easy</th><th>Moderate</th><th>Difficult</th><th>Very difficult</th></tr>
<tr><td>Simple past tense</td><td>When a finished action or event in the past is being expressed, the simple past tense is required.</td><td>He visited his brother yesterday.</td><td>*When he finished his homework, he <u>watch</u> a movie.</td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

The difficulty judgement questionnaires for applied linguists was based on the R&GG taxonomy.

The test of implicit L2 knowledge was an elicited imitation test (R. Ellis, 2006; R. Ellis et al., 2009; Erlam, 2006) comprising 78 sentences, with three grammatical and three ungrammatical sentences targeting each of the 13 grammar points, presented auditorily in a fixed pseudo-random order. Participants were required to listen to a sentence, make a truth judgement on the basis of their world knowledge and beliefs by ticking “true”, “not true” or “not sure” on an answer sheet, and were then required to orally repeat in correct English the sentence they heard.

The test of explicit L2 knowledge was a metalinguistic knowledge test consisting of two parts that comprised 35 items in total. Part 1 of the test was modelled on the instrument used by Ziętek and Roehr (2011) and required learners to correct highlighted errors at sentence level and provide the underlying pedagogical grammar rule in either English or Spanish. Part 2 of the test presented the targeted pedagogical grammar rules in both English and Spanish and asked learners to write correct English sentences fully illustrating each rule (Absi, 2014; Scheffler, 2011).

The results indicated that learners' and teachers' difficulty judgements are significantly correlated (*rho* = .63, *p* = .02), and teachers' judgements are correlated with the applied linguists' judgements of overall learning difficulty (*rho* = .57, *p* = .04). With regard to the relationship between perceived learning difficulty and learners' performance on both measures, only one significant correlation was found between learners' judgements and their scores on the explicit measure (*rho* = -.64, *p* = .02). Even though there is no significant correlation between teachers' difficulty judgements and learners' performance on either, this relationship approached significance for the elicited imitation test (*rho* = -.53, *p* = .06) and for the metalinguistic knowledge test (*rho* = -.52, *p* = .07).

In terms of the learners' performance on the measures of implicit and explicit knowledge, one finding that is shared between Ellis' (2006) study and Rodríguez Silva and Roehr-Brackin's (2016) study is that the grammar points that were discovered to be easy for learners in terms of their explicit knowledge were difficult in relation to their implicit knowledge and vice versa. Regarding the perceived difficulty of the targeted grammar points of applied linguists, learners and teachers, the results show that experts' learning difficulty judgements did not lead to significant prediction, and that the learners (*n* = 30) themselves rightly determined the learning difficulty for the 13 targeted grammar points with respect to the explicit measure, whereas the teachers' (n = 11) judgements showed a trend towards significance for both the test of implicit and the test of explicit L2 knowledge.

Like the other two approaches aforementioned (i.e. Spada & Tomita's (2010) and Roehr & Gánem-Gutierrez' (2009)), the approach of difficulty judgments of L2 experts presents its own problems. One problem that should be acknowledged is that the inclusion of experienced instructors' ratings of difficulty based on what grammar points are easy or difficult for their learners is that they rate the learning difficulty of structures in terms of group average and not for each individual. Although DeKeyser (2003) argues that such an approach may leave behind individual differences such as language aptitude (DeKeyser, 2003) and working memory (see sections 2.3 and 2.4 below), it can be argued that these two constructs may further explain why some learners find a specific grammar point difficult to learn while others may find it easy to learn.

Notwithstanding the limitations of the holistic judgements of L2 experts to determine the difficulty of grammar points, it is one viable approach as indicated by the empirical evidence in the studies previously discussed (Robinson, 1996; Rodríguez Silva & Roehr-Brackin, 2016; Scheffler, 2011). That is, given the different approaches (De Graaff, 1997; Goldschneider & DeKeyser, 2001; R. Ellis, 2006; Housen et al., 2005; Robinson, 1996) in determining the ease or difficulty of grammar points, using the holistic L2 experts' judgements as a subjective measure involves the elements of practicality (easy to administer and nonintrusive), cost (inexpensive), and high face validity (Révész, Michel, & Gilabert, 2016). In addition, teachers' education and experience as L2 learners (i.e., the cases where teachers had to learn the target language they teach) and as teachers qualify them as "well-suited candidates for empirically examining the easy-difficult distinction" (Graus & Coppen, 2015, p. 102).

It can be argued, based on the elements of practicality, cost, and face validity that holistic expert judgements is likely to be the most fruitful in determining the ease or difficulty of grammar points according to expert judgements of experienced L2 teachers. In addition to difficulty judgements based on group averages, it is important to consider cognitive IDs such as language aptitude and working memory capacity.

2.2.4 Interim summary

In this section it was discussed that no consensus has been yet reached in the research literature with regard to what makes the acquisition of different grammar points more or less difficult. Grammar points can be viewed as easy or difficult depending on a narrow or broader perspective (DeKeyser, 2005). That is, the form, function, and form-function mappings play a role in these two perspectives.

It was also discussed that even when specific criteria may depict the learning difficulty of grammar points as implicit and explicit knowledge (R. Ellis, 2006; DeKeyser, 2005), it is unclear how such criteria can be used to determine the relative learning difficulty of grammar points. It appears that R&GG taxonomy is the only taxonomy of its kind that can categorize learners' implicit and explicit knowledge of grammar points, but the results of a number of empirical studies (Huang, 2012; Rodríguez Silva & Roehr-Brackin, 2016; Thepseenu & Roehr, 2013; Ziętek & Roehr, 2011) are inconclusive.

It was suggested that an alternative approach in determining the ease or difficulty of grammar points is the holistic judgements of L2 experts (teachers). The findings of several empirical studies (Graus & Coppen, 2015; Robinson, 1996; Scheffler, 2011) lend support to this approach as one viable option to determine the ease or difficulty of grammar points.

However, by excluding the holistic judgements of learners leaves out an integrative element that can complement teachers' judgements and provide a fuller representation of the classification of the difficulty of grammar points if these two types of judgements are combined. Absi (2014) and Rodríguez Silva and Roehr-Brackin, (2016) are two examples that instantiate such an approach; in each of the studies learners scored the lowest on difficult grammar points indicating that taking into account both learners' and teachers' classification can be a fruitful approach to determine the learning difficulty of grammar points.

2.3 Language learning aptitude

In the previous section the variation in L2 learning among adult learners was discussed in terms of the implicit and explicit learning difficulty of grammar points. Such variation on L2 learning success can be further understood by analyzing learner characteristics—usually called individual differences (IDs) (Dörnyei 2005; Ehrman & Oxford, 1995). Dörnyei (2005) defines IDs such as language aptitude and working memory as "characteristics or traits in respect of which individuals may be shown to differ from each other" (p. 181).

Among the existing IDs such as intelligence, strategies, attitudes, anxiety, risk taking, introversion/extroversion, cognitive style, and ego permeability, language aptitude has been documented as one of the best predictors of L2 language learning success (DeKeyser, 2000; Erham & Oxford, 1995; Sawyer & Ranta, 2001; Skehan, 1989) both in global-domain area such as classroom contexts (Harley & Hart, 1997, 2002; Kormos & Sáfár, 2008; Wesche, 1981) and in specific-domain area such as the relationship to the learning of different grammar points (Robinson, 2002; Yalçin & Spada, 2016).

2.3.1 A historical overview of language aptitude

Language aptitude testing first became the object of interest in the 1920s and 1930s in the United States to identify students who had a low L2 performance in state schools. This was tackled with the administration of so-called “prognosis tests” to identify the causes of such failure (Spolsky, 1995, cited in Dörnyei, 2005). Thirty years later, language aptitude became, for the second time, the object of widespread interest and serious research in the 1950s with the design of the Modern Language Aptitude Test (MLAT) (Carroll & Sapon, 1959) and the Pimsleur Language Aptitude Battery (PLAB) (Pimsleur, 1966). Two decades later, in the 1970s and 1980s the position of language learning aptitude was not favourable in SLA research. This decreased interest in language learning aptitude was in part due to (a) language learning aptitude being associated with outmoded methodologies (Dörnyei, 2005; Ranta, 2008; Robinson, 2002; Skehan, 2002); and (b) the belief that instruction might not be effective for low aptitude learners, (c) and in the teaching profession of foreign languages there was very little interest in the differences that existed between learners (Skehan, 2002). Ten years later, researchers showed a renewed interest in the study of language aptitude (Dörnyei, 2005; Skehan, 2012) principally due to the developments in areas of cognitive psychology, such as working memory, as well as by developments in understanding how second languages are learned (Long & Doughty, 2009). Hence, it has been proved that language aptitude does not necessarily depend on any specific teaching method and that it is a good predictor of language learning success in both implicit and explicit conditions (Dörnyei, 2005).

The role the ID language aptitude plays nowadays in language research and language teaching is prominent, “it is now changing from a marginal position to one where it is centre-stage” (Skehan, 2012: 381). This claim is evidenced by theoretical perspectives (Ortega, 2009; Ranta, 2008; Skehan, 2012), and recent publications on the topic (Abrahamsson & Hyltenstam, 2008; Bylund, Abrahamsson, & Hyltenstam, 2010; Forsberg & Sandgren, 2013; Granena, 2013, 2014; Li, 2014; Linck et al., 2013; Sáfár & Kormos, 2008; Yalçin & Spada, 2016; Yilmaz, 2012).

2.3.2 Classic models of language aptitude

As previously mentioned, in the late 1950s and 1960s two commercial aptitude batteries were designed for different audiences: the MLAT (Carroll & Sapon, 1959) for adults, and the PLAB (Pimsleur, 1966) for adolescents. These two aptitude tests were developed to predict rate and success of classroom foreign language learning (Dörnyei, 2005; Ranta, 2008; Robinson, 2005) and to predict some areas of difficulty in the early stages of SLA such as phonological problems and syntactical difficulties (Grigorenko, Sternberg, & Ehrman, 2002). Particularly, the MLAT has been widely used in SLA research (see Li's, 2015, meta-analysis for a review).

According to Carroll's (1981: 105) aptitude model, language aptitude is composed of four constituent abilities:

- phonetic coding ability, defined as "the ability to identify distinct sounds, to form associations between these sounds and symbols representing them, and to retain these associations."
- grammatical sensitivity, which is "the ability to recognize the grammatical functions of words (or other linguistic entities) in sentence structures."
- rote learning ability/associative memory, which is "the ability to learn associations between sounds and meaning rapidly and efficiently and to retain these associations."
- inductive language learning ability, which is "the ability to infer or induce the rules governing a set of language materials, given samples of language materials that permit such inferences."

The MLAT and PLAB are language aptitude tests associated with this model. The former is composed of five parts: (a) number learning, (b) phonetic script, and (c) spelling clues, which measure the component phonetic coding ability; (d) words in sentences, which measures the component grammatical sensitivity; and (e) paired associates, which measures the component rote learning/associative memory. The latter language aptitude test is composed of six parts: (a) grade point average, (b) interest in foreign language learning, (c) vocabulary, (d) language analysis, (e) sound discrimination, and (f) sound-symbol association. As a result of the extensive use SLA researchers have made of these two aptitude tests, the MLAT, in particular, has remained virtually unchanged, though the conceptualization of language aptitude has undergone some development.

Some researchers have updated Carroll's classic model of language learning aptitude. Skehan (1998), for instance, has adopted Carroll's model to an information processing model; he developed a three-component model which comprises three constituent abilities, that is, phonetic coding ability, language-analytic ability, and memory. According to this aptitude model, there are some similarities and differences in comparison to Carroll's model. The constituent auditory ability and memory are similar to Carroll's phonetic coding ability and rote learning ability/associative memory. The main difference lies in the constituent language-analytic ability, which comprises Carroll's grammatical sensitivity and inductive language learning ability.

2.3.3 Recent developments of language aptitude measurement

The advances in cognitive psychology and the possible connections between language aptitude and other areas in SLA such as instructional treatment (Erlam, 2005; Kormos & Sáfár, 2008; Robinson, 2005; Yilmaz, 2012; Wesche, 1981), and structures (Robinson, 1996, 2002; Yalçin & Spada, 2016) have led to a revival of this individual differences variable in the past 25 years (Dörnyei, 2005).

The renewed interest in aptitude has also led to the design of language aptitude tests that can be widely available to SLA researchers and that can be easily administered through the use of computers (Granena, 2013). One of these language tests that comply with these parameters is the LLAMA test (Meara, 2005).

2.3.3.1 The LLAMA aptitude test

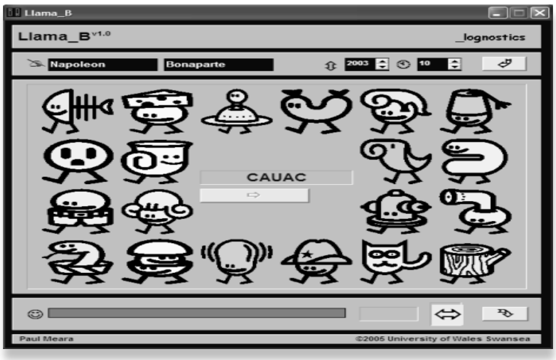
The LLAMA aptitude test (Meara, 2005) is a new version of the LAT (Meara, Milton, & Lorenzo-Dus, 2001). The LAT comprises five subtests. The LAT_A is a self-assessment test of aural memory for sound strings based on the Swahili language; the LAT_B is a test of visual memory of paired associates; the LAT_C is a test of people's ability to infer the rules of the artificial language "Novish"; the LAT_D is a test of a person's ability to recognize unfamiliar words; and the LAT_E is a test of a learner's ability to make connections between unfamiliar sounds and symbols; the tests are loosely based on the MLAT (Carroll & Sapon, 1959). The test received much interest from the research community and it was adapted to Swedish, Hungarian, and French versions, languages that use the Roman alphabet. The test was also requested to be adapted to Japanese, Greek, Russian, and Georgian languages, which do not use the Roman alphabet. The adaptation of the LAT test into these languages was not very successful because of failures to rework the program code to cope with these languages. Meara (2005) also reports that one of the tests (LAT_D) containing materials loosely based on Polish and Turkish turned out to be familiar to potential L1 Hungarian or Azeri test takers who probably had studied or spoke these languages. With these problems in mind, the authors of the LAT decided to develop the LLAMA test.

The LLAMA test is loosely based on the MLAT and is composed of sub-tests on paired associates learning, on probing sound recognition which requires previously heard sound sequences to be identified in new sequences, on targeting sound-symbol association, and on assessing grammatical inferencing. The authors of the LLAMA tests made the following modifications to each test: LLAMA_B is a new version of the LAT_B tests, which is based on picture stimuli, rather than the verbal stimuli of the original LAT_B; LLAMA_F is a new version of the LAT_C, which is also based on picture stimuli and has eliminated the need for an L1 data-base; LLAMA_D is a new version of the original LAT_D tests. The 2002 version of this test used stimuli loosely based on Turkish, whereas the new version uses stimuli based on a dialect of an indigenous language spoken in Northern Canada. A language aptitude test like this, which is based on picture stimuli and a dialect spoken in Northern Canada, can be used in different L1 contexts without granting advantages to any particular group, with the exception of those people speaking the dialect in Northern Canada. Each of these sub-tests is discussed in the sections that follow.

2.3.3.2 LLAMA B: Paired associates learning

This subtest (see Figure 2.1) measures the ability to learn new words. It presents 20 words associated to target images and is loosely based on the original vocabulary learning subtask of Carrol and Sapon (1959), the paired associates test.

Figure 2.1: LLAMA_B



There is a timed study phase (two minutes by default) in which test-takers click on the different images displayed on the screen. The name of each object is shown in the centre of the panel. In the testing phase, the program displays the name of an object and test-takers have to identify the correct image on the screen.

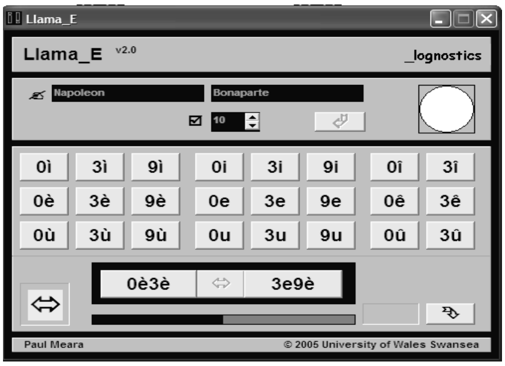
2.3.3.3 LLAMA D: Sound recognition

This subtest measures the ability to recognize patterns in spoken language. It does not include a timed study phase and presents a new task that does not appear in the work of Carroll and Sapon (1959). In this test, test-takers listen to a string of 10 sound sequences only once that are computer-generated and based on the names of objects in a British-Columbian indigenous language. Then, test-takers complete a recognition test where they have to identify whether they heard a sound sequence in the presentation phase or not.

2.3.3.4 LLAMA E: Sound-symbol association

This subtest (see Figure 2.2) measures the ability to form novel sound-symbol associations (phonetic coding ability).

Figure 2.2: LLAMA_E



The test requires test-takers to work out relationships between 24 sounds (i.e. recorded syllables) and a written representation of the syllables in an unfamiliar alphabet. The test includes a timed study phase (two minutes by default) in which test-takers click on the different symbols displayed and try to learn the corresponding sound association. They then hear a combination of two syllables and have to decide its symbol correspondence by clicking on the right button.

2.3.3.5 LLAMA F: Grammatical inferencing

This subtest (see Figure 2.3) measures the ability to infer or induce the rules of an unknown language (i.e. language-analytic ability).

Figure 2.3: LLAMA_F

The test shows 20 pictures and sentences describing the pictures. Test-takers have to work out the grammatical rules that operate in the language. The test includes a timed study phase (five minutes by default) in which test-takers click on a series of small buttons displayed on the screen. For each button, a picture and a sentence describing the picture are displayed. In the testing phase, the program shows a picture and two sentences, one grammatical and one ungrammatical, and test takers click on the one they think is correct. While the LLAMA_D and LLAMA_E subtests can be assumed to reflect the efficiency of creating phonological representations in working memory, the LLAMA_B and LLAMA_F subtests involve processing of verbal material and thus the manipulation of information in working memory.

2.3.3.6 The CANAL-F test

Other recent language aptitude tests such as the CANAL-F test (Cognitive Ability for Novelty in Acquisition of Language—Foreign) were developed by Grigorenko, Sternberg, and Ehrman (2002). This language aptitude test was designed for learners to cope with novelty and ambiguity. The theory of the test underlines the mental processes of selective encoding, accidental encoding, selective comparison, selective transfer, and selective combination. Its sampling frame includes concern for levels of processing, modes of input, and encoding, storage and retrieval of information, raising the need to have immediate and delayed recall. The test is based on the invented language Ursulu and is an integrated test (i.e. the different subtests are cumulative in what they assess), and it only exists for L1 English speakers and takes several sessions to administer. There are five sections: learning meanings of neologisms from context, understanding the meaning of passages, continuous paired associates learning, sentential inference, and learning language rules.

2.3.4 Differences between language aptitude tests

If a comparison is made between the language aptitude tests described above (i.e. MLAT, PLAB, LLAMA and CANAL-F), it can be seen that each aptitude test focuses on specific factors, as shown in Table 2.6.

Table 2.6 A comparison of different foreign language aptitude batteries

	MLAT	PLAB	LLAMA	CANAL-F
Phonetic coding ability	+	+	+	
Grammatical sensitivity	+	+		
Inductive language learning		+	+	+
Paired associates	+		+	+
Working memory			+	
Attentional processing				+
Working memory to long-term memory connections				+

(Adapted from Skehan, 2012, p. 390)

Table 2.6 shows that the MLAT omits inductive language learning ability, which is part of Carroll's (1981) aptitude model. Another ability that is absent from Carroll's model is working memory (Baddeley, 2007). This is understandable taking into account that working memory was not the type of memory dominant in psychology at the time of development of the MLAT (Skehan, 2002), that is, Carroll developed a section in the MLAT on associative memory because that was the type of memory associated with instructional contexts at that time. Unlike the MLAT test, the PLAB and CANAL-F test do include inductive language learning ability. The CANAL-F also includes attentional processing ability (i.e. processes of selective encoding, accidental encoding, selective comparison, selective transfer, selective combination) and working memory to long-term memory connections ability.

In contrast to the MLAT, PLAB and CANAL-F, the LLAMA test includes working memory ability in the subtests LLAMA_D and LLAMA_F. This sole difference makes the LLAMA test attractive in SLA research even more so because nowadays the MLAT is only available to government agencies, missionary groups, and licensed clinical psychologists (http://lltf.net/aptitude-tests/language-aptitude-tests/modern-language-aptitude-test-2/), the PLAB is intended for younger populations, and the CANAL-F has a military history and is restricted (Skehan, 2012), and it only exists for L1 English speakers and takes several sessions to administer. Furthermore, the LLAMA test "adds a receptive interpretation of inductive language ability" (Skehan, 2012, p. 390). This means that the LLAMA test provides an opportunity for learners to use strategies and problem-solving techniques; this is the case for LLAMA_F which instructs learners to work out the relationship between grammatical rules and pictures. On LLAMA_B, even though learners are not explicitly instructed to work out the relationships between objects and names as a strategy (Granena, 2013), they may opt to do so to complete the task. LLAMA_D is the only subtest that does not include a study phase, and therefore, does not allow time to rehearse; Granena (2013) argues that this characteristic minimizes problem-solving and strategy. Unlike the other subtests, LLAMA D requires implicit cognitive and memory processes whereas LLAMA B, E, and F require explicit cognitive and memory processes.

Bearing these characteristics of the LLAMA test in mind, its availability to SLA researchers, its relatively easy administration, and its language independence, the LLAMA test is a viable option to gauge language learning aptitude in learners.

2.3.5 Validation of the LLAMA test

Even though the LLAMA test has been used in a number of studies in SLA research (Cherciov, 2011; Forsberg & Sandgren, 2013; Granena, 2012; Yalçin & Spada, 2016; Yilmaz, 2012), the author (Meara, 2005) acknowledges that the LLAMA test has not been standardized. Nevertheless, the validity and reliability of the test was assessed in two studies (Granena, 2013; Rogers et al., 2016), and the test has proved to be valid given the expected results in a number of studies (Cherciov, 2011; Forsberg & Sandgren, 2013; Granena, 2013; Yilmaz, 2012).

In an exploratory validation study (Granena, 2013), with the aim to assess the reliability of the LLAMA in terms of its internal consistency and stability in time, a total of 186 participants from three different L1 backgrounds (Chinese, Spanish, and English) were included in the study. All participants completed the four LLAMA subtests in a random order, and their performance was audio-recorded, that is, the program gives test-takers feedback in the form of a *ding* for a correct answer, and a *bleep* for a wrong answer. These two sounds were audio-recorded when participants were completing each subtest and later transcribed into the binary code (1 and 0) needed to compute internal consistency. The LLAMA test does not provide itemized data and recording every answer is a solution to compute reliability on this test. The results indicate that each of the four LLAMA subtests had acceptable reliability on the two different indices measuring internal consistency and stability over a period of two years.

In the second validation study (Rogers et al., 2016), the researchers investigated a number of areas pertaining to the LLAMA tests such as (1) the role of gender in LLAMA test performance, (2) language neutrality, (3) the role of age, (4) the role of formal education qualifications, (5) the effect of playing logic puzzles on LLAMA scores, and (6) the effect of changing the test timings. Two hundred and twenty nine participants from a range of language backgrounds, aged 10-75 with various education levels, typologically distinct L1s, and varying levels of multilingualism took part in the study. A subset of 65 participants was also tested with varying timings for the tests. The following results were obtained: (1) no significant differences between male and female participants for any of the LLAMA tests; (2) no significant differences between monolingual, bilingual, and multilingual groups; (3) the younger learners (aged 10-11) performed significantly worse than the adults in the sound/symbol correspondence task (LLAMA_E); (4) Formal education qualifications showed a significant advantage on LLAMA_B, LLAMA_E, and LLAMA_F but not on LLAMA_D; (5) playing logic puzzles did not improve LLAMA test scores; and (6) the timings appear to be optimal for LLAMA_B, LLAMA_D, and LLAMA_E except for LLAMA_F for which a decrease of time could be considered.

2.3.6 Empirical research on aptitude using the LLAMA test

A number of studies have used the LLAMA test to measure learners' language learning aptitude. For instance, in Cherciov's (2011) study about L1 attrition (Romanian), a bilingual group of participants (*N* = 20) of Romanian speakers residing in Canada were administered the LLAMA test to see to what extent their performance on this test would predict not only L1 maintenance, but also L2 achievement. The L1 and L2 proficiency of the participants were measured through a C-test, a verbal fluency test, and a spontaneous speech production task in the form of a film commenting task. The results indicate that high LLAMA scores did not appear to influence L1 maintenance (or the avoidance of L1 attrition). However, high levels of L2 acquisition correlated with high language aptitude.

In another study (Forsberg & Sandgren, 2013) that recruited 13 L1 Swedish highly proficient learners of L2 French who completed a GJT, a collocation test, and the LLAMA test, only one significant positive correlation was reported between the LLAMA_D subtest and the collocation test scores. However, trends in the direction of significance were found between LLAMA_F and GJT scores, and between LLAMA_B and LLAMA_F.

In one more study (Yilmaz, 2012) forty-eight participants were randomly assigned to a recast group (*n* = 16), an explicit correction group (*n* = 16), and a control group (*n* = 16). Each learner was asked to complete three measures: an oral production test in which learners were asked to describe pictures; a comprehension test where learners were asked to choose the picture that illustrated the verbal stimuli, and a recognition test in which learners had to choose the correct word among four words that would complete a sentence. The two target structures selected for the study were the Turkish plural morpheme /-lAr/, and the Turkish locative case morpheme /-DA/. The results indicated a strong, statistically significant relationship between post-test performance and grammatical inferencing (LLAMA_F) under the explicit feedback condition (i.e. explicit correction) among L2 learners with high LLAMA_F scores, but no significant relationship was found for the implicit feedback condition (i.e. recasts).

Apart from these areas of research, other areas of research interest of language aptitude to predict L2 learning outcomes in a range of environments and learning conditions (Roehr, 2012) have been explored in SLA.

2.3.7 An overview of the role of aptitude in SLA

Li's (2015) meta-analysis provides an overview of research on aptitude. This study was guided by the researcher's interest in (a) the relationship between language aptitude and other individual difference variables such as motivation, anxiety, and intelligence, (b) the relationship between aptitude and working memory, and (c) the relationship between aptitude/aptitude components and general L2 proficiency, specific aspects of L2 learning such as L2 knowledge (grammar and vocabulary) and L2 skills (listening, speaking, reading and writing). The whole dataset consisted of 66 studies that reported data contributed by 109 unique samples. Among the 66 retrieved studies, 47 are published, and 19 are PhD and MA dissertations. The aptitude tests included in the study are the MLAT and its adapted versions in other languages, VORD, LLAMA, and the Language Analysis Test (a measure of language analytic ability). In terms of instructional context, most of the studies were carried out in high school (*N* = 22) or university (*N* = 26) language programs, and only a small number of studies were conducted in other contexts. The majority of the studies (*N* = 49) were conducted with L1 English speakers who were learners of French (*N* = 24), Spanish (*N* = 20), and German (*N* = 12), and the studies were primarily undertaken in foreign language setting (*N* = 53) where the target language is not the language spoken in the community, rather than in second language settings (*N* = 10) in which the target language is also the language of the community. The study obtained the following findings: (1) aptitude is independent of other cognitive and affective factors such as motivation, anxiety and intelligence, (2) executive working memory was more strongly associated with aptitude and aptitude components than phonological short-term memory, (3) aptitude measured using full-length tests was a strong predictor of general L2 proficiency, but it had low predictive validity for vocabulary learning and L2 writing, and (4) different aptitude components demonstrated differential predictive validity for different aspects of learning. In this last result, the aptitude component language analytic ability was the strongest predictor of grammar learning, the aptitude component phonetic coding was the weakest predictor for listening comprehension, and the language aptitude rote memory/associative memory was almost invariably the weakest predictor among all aptitude components/subtests.

Another area of research interest of language aptitude focusing on processes rather than products (DeKeyser, 2012) is emerging in recent research. That is, the interaction between aptitudes and treatments point to different learning processes (DeKeyser, 2012).

2.3.8 Aptitude-treatment interaction research

A number of studies on language aptitude has been concerned with whether there is a relationship between aptitude and L2 learning in different instructional settings.

The interaction of aptitude differences with different types of language instruction has been demonstrated by Wesche’s (1981) seminal study. Wesche investigated the French language training program of the Public Service Commission of Canada, where language aptitude tests, that is, a combination of the five subtests of the MLAT and the Sound Discrimination and Sound Symbol Association subtests of the PLAB were given to candidates for prognostic and diagnostic purposes before entering the program. The program offered three different types of language instruction: (a) an audio-visual method, (b) an analytical approach, and (c) a functional approach. Most learners were placed in the core method (the audio-visual method). Those learners who had good scores on the Words in Sentences and Spelling Clues subtests of the MLAT were assigned to the analytical approach. Those learners who had good memory and auditory abilities as reflected on the Number Learning, Phonetic Script and Paired Associates of the MLAT, and Sound Discrimination and Sound-Symbol Association of the PLAB but achieved low scores on the tests measuring analytical abilities were placed in the functional approach. The results show that the learners matched to the method according to their aptitude profiles reported overall satisfaction with the methods assigned to them. A further verification was conducted to check whether there was any difference in performance in language activities between appropriately matched students and mismatched students, that is, half of the students of the group of the analytical approach were placed in the audio-visual approach. The students who were placed in the right type of class according to their profiles achieved superior scores on three of the four achievement measures of listening comprehension and oral expression.

Likewise, whether aptitude is relevant in communicative language teaching is a controversial issue (Krashen, 1981; Ranta, 2008; Cook, 1996; Carroll & Sapon, 1959). One study that lends support to Carroll and Sapon's proposition that the MLAT addresses learning abilities that are independent of methodology is that of Ehrman and Oxford (1995). This study consists of a sample of 855 people of which 282 completed the MLAT. These researchers found a relationship between the MLAT scores and L2 proficiency scores in their large-scale study of US government employees. The method of instruction was largely communicative. They found a moderate correlation between the Words in Sentences subtest of the MLAT and learners' speaking and reading scores.

Further empirical evidence that aptitude is relevant not only in communicative language teaching but also in an immersion program is Ranta's (2002) study. The emphasis of the study was on developing communicative skills and the learners received no form-focused instruction. Students (N = 150) in the study attended a school where they received five months of intensive English a second language (ESL) instruction in grade 6. Learners had five classes with an all-day program in ESL. An L1 (i.e. French) metalinguistic task was designed to measure learners' language analytic ability based on error detection and correction task. The L2 proficiency measures used were an aural vocabulary recognition test, a listening comprehension test, a cloze test, a Ministry of Education of Quebec listening comprehension test, an English metalinguistic test, and a yes/no vocabulary test. The participants were tested on three occasions during their intensive ESL year. The researcher found significant correlations between the L1 metalinguistic task and learners' performance on the L2 measures, though the interpretation of these results can be questioned in the sense that the researcher did not use an existing test of aptitude but one designed on her own.

In a relatively recent study, Erlam (2005) investigated whether there was any relationship between the effectiveness of three different instructional methods (i.e., deductive instruction, inductive instruction and structured input instruction) and language aptitude and working memory, as measured by a self-designed test, multisyllabic word test. Second year high school learners of L2 French (*N* = 92, approximately 14 years of age) were assigned to three different instructional groups. Each group received 3 x 45 minutes of instruction with the researcher. The target structure was direct object pronouns in French. All students were assessed on measures of listening comprehension, reading comprehension, written production and oral production, over three testing sessions (i.e., pre-test/post-test/delayed post-test). Six months after the completion of the study, two language aptitude tests (i.e., the Sound Discrimination test of the PLAB (Pimsleur, 1966), and the Words and Sentences subtest of the MLAT (Carroll & Sapon, 1959), and the phonological short-term memory test were administered. The results indicate that the students in the deductive instruction group made greater gains than the students in the inductive instruction group and this latter group performed better than the students in the structured input instruction group. The results further suggest that for the most part, differences in individual gains did not correlate with differences in language aptitude for the deductive instruction group of learners. Thus, it seems that deductive instruction may neutralize individual differences in language aptitude. In contrast, students who had greater language analytical ability as measured by the Words and Sentences subtest of the MLAT gained more from inductive instruction, followed by the structured input group, as evidenced on the written production delayed post-test (r = 0.59 and r = 0.51 respectively). Another interesting result in the inductive instruction group is that greater analytical ability correlated negatively with production of direct object pronoun forms on the oral test (r = -0.61), which could be interpreted that students with greater analytical ability had a low performance on the target structure in a test where there was greater time pressure and where a focus on meaning was needed. With respect to the deductive instruction group and structured input instruction group, learners with strong language analytic ability gained more from these types of instruction than the other learners.

As for the variable working memory, the results show that the learners with greater working memory capacity benefited more from structured input instruction, in terms of production of the target structure. Erlam (2005) argues that “the higher correlations for delayed rather than immediate post-test gain scores is evidence that learners who had greater working memory processing capacity were more successful at maintaining long-term representations of the target language forms” (p. 166).

2.3.9 The relationship between aptitude and learning difficulty

Regarding the relationship between aptitude and learning difficulty, two studies (Robinson, 1997; Yalçin & Spada, 2016) have provided evidence that there is a relationship between language aptitude and learners’ performance on easy and difficult grammar points.

In Robinson’s (1997) study, as mentioned in Section 2.2.3, the following factors were included:

1. Learning condition, with four levels: implicit condition, incidental condition, rule-search condition, and instructed condition;

2. Rules to be learned based on two criteria – the relative complexity of the structures described by the pedagogic rules, and the relative complexity of the pedagogic rules describing the structures (a simple rule describing the fact that subject-verb inversion is allowed in sentences where adverbials of movement/location are fronted, e.g., “Into the house John ran/ran John”; and complex rule describing how to form pseudo-clefts of location, e.g., “Where Mary and John live is in Chicago not in New York”);
3. Two aptitude components: Memory (the MLAT paired-associate test) and Grammatical sensitivity (the MLAT words in sentences test);
4. Awareness: Debriefing questionnaire (did participants notice any rules? were they looking for rules? could they verbalize rules?)
5. Learning was evaluated by a grammaticality judgement task.

The sample included 104 learners of English as an L2, most of whom had Japanese as their L1. The learners were pretested on their knowledge of the target structures (i.e. subject-verb inversion and pseudo-clefts of location) and those who showed by their performance on the pre-tests that they were unfamiliar with those structures were selected for the study. The participants were randomly assigned to one of four learning conditions: implicit, incidental, rule-search, and instructed conditions. Sentences were presented to all the participants on a computer for the same length of time and according to each condition participants performed different tasks: in the implicit condition the participants were told that they would perform a memory task, and they were asked questions about the location of words in sentences; in the incidental condition the participants were asked to answer yes/no comprehension questions, and they received feedback about their responses; in the rule-search condition the participants were told to try and find the rules underlying the sentences, and they were asked to state if they had found the rules or were still looking and no feedback was provided; and in the instructed condition the participants were given explanations of the easy and hard rules. These explanations were followed by questions about the metalinguistic form of the sentences, such as “Did the subject of the sentence come before the verb?” Participants responded yes or no to each question and received feedback.

The results show that the instructed group outperformed all the other groups with respect to the easy and hard rule. Further results showed that easy rule sentences were predicted by grammatical sensitivity but not by memory and difficult rule sentences were predicted by memory but not by grammatical sensitivity, and this result was obtained in the rule-search condition only (i.e. inductive explicit learning). The strongest correlations of performance on the aptitude subtests were with accuracy on the easy and difficult rules in the implicit condition. According to the debriefing questionnaire, the learners in the implicit condition who responded that they had looked for rules, and those who were able to verbalize rules were more accurate on the grammaticality judgement task. Furthermore, those learners who were significantly better on the MLAT words in sentences subtest were the ones who looked for rules or could verbalize the rules.

Robinson’s (1997) interpretation of these results is that the instructed group outperforming the other groups on the easy and hard rule is probably short-lived, explaining that “learning resulting from other conditions, though far less extensive, may be more permanent” (p. 83). Robinson (1997) also interprets the finding on easy rule sentences being predicted by grammatical sensitivity but not by memory, and difficult rule sentences by memory but not by grammatical sensitivity as a reflection of the processing operations of distinct unconsciously and consciously accessed systems (i.e. memory and grammatical sensitivity respectively). One more interpretation pertaining to the result of the implicit group of participants is that participants with high aptitude “became aware of the rule-governed nature of the sentences during training and began to actively search for rules” (Robinson, 1997, p. 82).

In a more recent study (Yalçin & Spada, 2016), a quasi-experimental pre-/post instructional intervention study, 66 pre-intermediate, secondary-level learners of L2 English enrolled in the eighth grade (13 to 14 years old) at a private secondary school in Turkey were pretested with two written and untimed GJT and two picture-cued oral production tasks before the instruction began. The four measures targeted two L2 structures which were classified as easy and difficult following the criteria of formal complexity, input frequency, salience differences and participants' perceived difficulty of the two grammar points: the passive (the "difficult" grammar point) and the past progressive tense (the "easy" grammar point). The two grammar points were taught for four hours each by one of the researchers, and after the instructional intervention, the measures were administered again. In addition, within the first week of instruction the LLAMA aptitude test was administered. The results indicate that the grammatical inferencing component (LLAMA_F) contributed to learners' gains on the passive but not on the past progressive on the written measure. On the other hand, the associative memory component (LLAMA_B) contributed to learners' gains on the past progressive on the oral measure.

The results of this study seem to indicate that learners may use different aptitude abilities depending on the degree of difficulty of the grammar point being measured, the length of exposure and the degree of novelty of the grammar point (Yalçin & Spada, 2016). The degree of novelty of a grammar point may depend on the L2 proficiency of the learner, that is, high-proficiency learners have more language learning experience than low-proficiency learners.

In the last two studies reported in this section (Robinson, 1997; Yalçin & Spada, 2016), the findings indicate that different aptitude components can aid learners in the acquisition of easy or difficult grammar points. This supports Skehan's (2002) proposal that different components of aptitude may be involved at different stages of L2 acquisition. Learners' development in SLA can be understood as the developmental stage in accordance with Skehan's (2002) aptitude profile model. This model suggests nine developmental stages with four main levels:

"Stage 1 Noticing: …the initial inroad, the first insight (within any particular domain) that some aspect of form is worth attention.

Stages 2-5 Patterning: …the capacity to detect and manipulate patterns in the target language.

Stages 6-8 Controlling: …the development of control in a process where a rule-based generalization, initially handled with difficulty, becomes proceduralised.

Stages 8-9 Lexicalizing: …the learner is able to go beyond rule-based processing, however fast, and build a lexical system which can be used to underlie real-time performance" (p. 90).

It can be hypothesized that some LLAMA subtests such as the associative memory component (LLAMA_B) and the sound recognition component (LLAMA_D) may predict learners' acquisition process in the acquisition of easy grammar points in the first two main levels. Likewise, the grammatical inferencing component (LLAMA_F) may predict learners' acquisition of difficult grammar points in the last two main levels. However, these hypotheses do not hold true in Robinson's (1997) study; his findings are the opposite, that is, the associative memory component predicted difficult rule sentences and grammatical sensitivity easy rule sentences. The opposing findings in both studies may be due to the different types of aptitude measures.

It can be noted that these two studies and the other studies discussed in this section do not address the role language aptitude plays in the learning of easy and difficult grammar points at different levels of proficiency (Yalçin & Spada, 2016; Hummel, 2007). Such an approach may indicate whether there are relevant interactions between language aptitude, easy and difficult grammar points, and level of proficiency.

2.3.10 Interim summary

In this section, the variation in L2 learning in adult learners, apart from differences in age, L1, overall proficiency and stage of development, and analytic mental processes, to mention a few, was in part attributed to their language learning aptitude, which is reflected in their explicit knowledge as well as their implicit knowledge of L2. The studies reviewed in this section demonstrate ample research on the role of aptitude in SLA, but there is scarce research conducted on the differentiation between implicit and explicit knowledge and learning (Robinson, 1997; Yalçin & Spada, 2016).

Different measures of language aptitude have been developed in the past five decades reflecting, to some extent, the changes in methodology in teaching foreign languages, that is, from the Audio-lingual Method focusing on oral mechanical drills such as repetition, restatement, integration (Richards & Rodgers, 1986) to the communicative approach focusing on using language for meaningful interaction and for accomplishing tasks (Lightbown & Spada, 2001). In addition to this, language aptitude measures have changed, particularly in memory orientation, from an associative memory where information is stored to working memory where information is stored and manipulated.

Aptitude research reveals that high levels of language aptitude correlate with high levels of L2 acquisition (Cherciov, 2011; Forsberg & Sandgren, 2013; Yilmaz, 2013) as well as different aptitude components demonstrate differential predictive validity for different aspects of learning such as grammar and listening comprehension (Li, 2015). It also reveals that in different learning contexts such as a communicative context (Ehrman & Oxford, 1995) and in immersion programs (Ranta, 2002), aptitude plays an important role as evidenced in the relationships between learners' aptitude and learners' scores on speaking, reading, listening, and metalinguistic measures. A further revelation is that language aptitude interacts with the type of instructional context (Erlam, 2005; Robinson, 1997; Wesche, 1981), and the type of grammar points (i.e. difficult grammar points vs. easy grammar points) included in research studies (Yalçin & Spada, 2016). With respect to the relationship between aptitude (components) and easy and difficult grammar points, conflicting results have been found (Robinson, 1997).

With regard to Erlam's (2005) findings, students who had greater language analytical ability as measured by the Words and Sentences of the MLAT gained more from inductive instruction and structured input instruction than from deductive instruction. Although, the researcher argues that a possible explanation could be that deductive instruction that gives students opportunities to produce language output may neutralize individual differences in language aptitude, it is difficult to infer from this study whether the variable aptitude may be important at low or high levels of proficiency because the learners' level of proficiency in French is not reported.

The different findings in the studies reviewed in this section suggest that high levels of L2 acquisition correlated with high language aptitude (Forsberg & Sandgren, 2013), different aptitude components such as language analytic ability correlated with difficult grammar points and associative memory ability with easy grammar points (Yalçin & Spada, 2016), and explicit instruction led to significant relationships between learners' performance on different written and oral measures and aptitude (Erlam, 2006; Robinson, 1997; Yilmaz, 2012). As convincing these results may seem, scrutiny of their research designs, in particular to the range of age of the participants, number of grammar points included in the studies, and the type of aptitude measure reveal that these are variables that influenced in the type of result obtained in each study. For instance, Erlam (2006) used the Sound Discrimination test from the MLAT, and the Words and Sentences subtest from the PLAB to measure learners' language aptitude; other studies used a self-designed test (Ranta, 2002), and other studies reported correlations with aptitude components but not with the overall score of the aptitude test (Ehrman & Oxford, 1995; Yalçin & Spada, 2016; Yilmaz, 2012). Another example is the similarity of the number of grammar points included in each study; none of the studies discussed above included more than two grammatical structures. Hence, the results of these studies may not show conclusive evidence of the role of aptitude in L2 learning, namely, in relation to implicit and explicit knowledge of easy and difficult of grammar points.

Furthermore, as already indicated in the description of the LLAMA test, working memory may be a component of language aptitude (Dornyei, 2005; Miyake & Friedman, 1998; Sawyer & Ranta, 2001). Skehan (2012) endorses this suggestion by considering the potential involvement of working memory in different stages of L2 processing such as input processing and noticing and handling form and meaning.

2.4 Working memory capacity

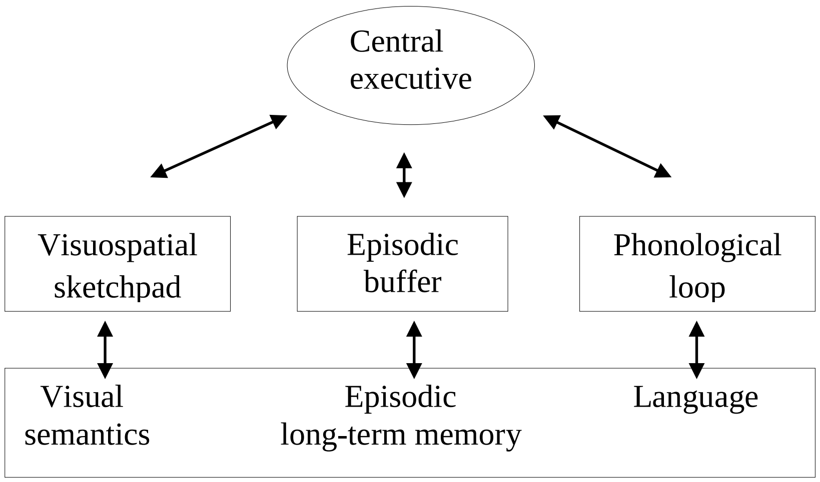
The ID working memory capacity is another predictor that plays a role in the development of L2 implicit and explicit knowledge (Erçetin & Alptekin, 2013; Sanz, Lin, Lado, Stafford, & Bowden, 2014; Serafini & Sanz, 2015).

Working memory is a cognitive ability that is responsible for “temporary storage and manipulation of information that is assumed to be necessary for a wide range of complex cognitive activities” (Baddeley, 2003, p. 189). In the literature, a number of working memory models have been put forward (Juffs & Harrington, 2011). The Just and Carpenter (1992) model is a unitary model of WM that comprises the integration of the storage and processing functions of working memory in language comprehension. Both of these functions are fuelled by the same commodity: activation. In this model, “capacity can be expressed as the maximum amount of activation available in working memory to support either of the two functions” (Just & Carpenter, 1992, p. 123). Hence, the WM capacity of an individual is determined by the amount of available activation. In the Ericsson and Kintsch (1995) model, the proponents view cognitive processes as a sequence of stable states that represent end products of processing. These end products of processing are stored in long-term memory, and they can be accessed by means of retrieval cues in short-term memory. The two most widely recognized WM models are the Just and Carpenter (1992) model and the Baddeley and Hitch (1974) model. The latter model has been one of the most influential models in SLA (Juffs & Harrington, 2011; Miyake & Friedman, 1998; Serafini & Sanz, 2015; Wen, Mota, & McNeill, 2015). Unlike the Just and Carpenter (1992) model which is a unitary model of WM, in the Baddeley and Hitch (1974) model working memory comprises multiple specialized components where cognitive processes occur in a coordinated manner (Baddeley & Logie, 1999). These specialized components of cognition are described in Baddeley and Hitch's (1974) original model and in its updated version (Baddeley, 2000).

2.4.1 Baddeley's multi-componential model of working memory

Baddeley and Hitch's (1974) and Baddeley's (2000) multi-componential model of working memory comprises a limited-resource central executive system, an episodic buffer, and two subsidiary "slave" systems: the visuospatial sketchpad and the phonological loop (see Figure 2.4).

Figure 2.4 Baddeley's (2000, p. 421) model of WM



The central executive is responsible for the coordination of the slave systems (i.e. the visuospatial sketchpad and the phonological loop), attentional control, manipulation of information, retrieval of information from long-term memory, and inhibition of information that might distract from or interfere with successful task execution. The visuospatial sketchpad is responsible for the storage and manipulation of visuospatial information. The phonological loop is responsible for the storage and manipulation of verbal material. It comprises a phonological store which can hold information for very brief periods of time (of around two seconds) and an articulatory rehearsal component which can be used for keeping the information active in the phonological store through subvocal articulation. The episodic buffer is in charge of integrating information to form episodes, and is in communication with long-term memory (Baddeley, 2000). L2 learning researchers have mainly focused on measuring the capacity of the phonological loop (i.e. phonological short-term memory) (Baddely, 2012) and to a lesser extent the central executive (i.e. complex working memory) (Hummel, 2009; Sáfár & Kormos, 2008; Skehan, 2012; Wen, 2012). In order to measure these two memory systems, researchers have developed complex span tasks to gauge complex working memory, and simple span tasks to gauge phonological short-term memory.

2.4.2 Measures of phonological short-term memory

Phonological short-term memory tasks such as word span, digit span, and non-word span have been developed to primarily tap storage of verbal information (Juffs & Harrington, 2011; Baddeley, 2003), which is a function of the phonological component of working memory. These tasks involve the recall of sets of unrelated words or numbers, presented in written or aural mode. Items are usually presented in sets of ascending size, and test-takers have to recall each set of items; if test-takers fail to correctly repeat two sequential sets, the task is terminated (Wechsler, 2003). The use of the word span test poses the plausible situation that test-takers could have prior knowledge of the word, thus confounding to some extent phonological short-term memory capacity (PSTM capacity) and language knowledge. To lessen this effect to some extent, researchers opt for using a digit span test (Harrington & Sawyer, 1992) or a non-word span test.

2.4.3 Measures of complex working memory

Unlike the measures of phonological short-term memory that primarily tap storage of verbal information, measures of complex working memory such as the reading span task and the listening span task (Daneman & Carpenter, 1980), the operation span task (Turner & Engle, 1989), and the backward digit span task (BDS task) (Kormos & Sáfár, 2008) aim at engaging both storage and processing capacity (Juffs & Harrington, 2011) by asking participants to temporarily hold information in memory (words, digits or letters) while carrying out a number of other cognitive operations at the same time. These complex span tasks better represent the conditions of online comprehension and production of language (Gilabert & Muñoz, 2010). One of the tasks to measure complex working memory is Daneman and Carpenter's (1980) reading span test. In this test participants are asked to read aloud increasingly longer sets of sentences, and at the end, participants need to recall the final words of all the sentences in each sequence. Another well-known measure of complex working memory capacity is the operation span task (Turner & Engle, 1989). In this test, simple arithmetic equations are employed instead of sentences, and participants need to solve mathematical operations while they have to remember words or letters that are paired with each equation. An alternative to these two complex span tasks is the backward digit span task (Kormos & Sáfár, 2008). This test lessens the demand on language knowledge (Juffs & Harrington, 2011) by using digits instead of words or sentences. This complex span task contains a series of spoken random digits which are presented to the participant who needs to retain the sequence of digits and then manipulate that information by repeating the numbers in reverse order. Unlike the simple span tasks which only place a cognitive demand for storage of information, complex span tasks tax the working memory capacity of a learner for storing, processing and manipulating information.

2.4.4 The relationship between WM capacity and language proficiency

While greater phonological short-term memory is important for L1 vocabulary learning and L2 word learning at early stages (Cheung, 1996; Gathercole & Baddeley, 1989; Juffs & Harrington, 2011; Wen, 2012), greater complex working memory is conceived as more relevant for L2 performance in a number of areas in language proficiency such as speech production, reading, writing and grammar (Gilabert & Muñoz, 2010; Harrington & Sawyer, 1992; Kormos & Sáfár, 2008; Linck & Weiss, 2011; Mizera, 2006; Roehr & Gánem-Gutiérrez, 2009; Sanz et al., 2014; Serafini & Sanz, 2015; Trebits & Kormos, 2008).

Harrington and Sawyer's (1992) study is one of the few early studies in SLA that included a measure of complex WM capacity and proficiency. The study was conducted at the International University of Japan. Thirty-four advanced learners studying in the university's Intensive English Program participated in the study. The study used the learners' TOEFL (Grammar and Reading sections) scores as a measure of their proficiency and the English reading span test as a measure of their complex WM capacity. This test consisted of 42 sentences that were simple, active, and 11 to 13 words in length. The sentences were 3 to 4 words shorter and simpler syntactically than those used for L1 speakers of English in Daneman and Carpenter's (1980) study in order to avoid possible floor effects in performance due to task difficulty. The results showed that the L2 English reading span measure had a significant correlation with the two TOEFL measures: L2 reading span and TOEFL Grammar ($r = .57$) and L2 reading span and TOEFL reading ($r = .54$).

In contrast to the research design of Harrington and Sawyer (1992), Mizera (2006) used three different WM tasks that 44 L1 English learners of Spanish completed in their L1: two complex measures of working memory, a speaking span task and a math span task; and a measure of phonological short-term memory, a non-word repetition task. Both types of measures were employed to explore the relationship between working memory and L2 proficiency. The learners' current level of attainment was measured by a general Spanish proficiency test. Mizera used a narrative monologue test, a comic strip task, an English-to-Spanish word translation test, and an imitation/grammaticality test to elicit L2 speech. The results showed weak correlations between all the measures of working memory and learners' performance on the L2 oral fluency tests (the narrative monologue test and the comic strip task). These same correlations became moderate when only considering the performance of the 20 most proficient students. These results were unexpected because it was hypothesized that strong correlations would be found between working memory capacity and L2 oral fluency. Mizera (2006) explains that these findings are probably due to "the complex nature of speaking fluently in a foreign language, which may call on faculties other than working memory, the persistence of working memory's influence over the course of language learning, and personal and affective factors" (p. 105).

In another study (Trebits & Kormos, 2008), in which a backward digit span was used, the researchers looked at the correlation between complex WM capacity and L2 fluency, complexity, and accuracy on the performance of both a simple and a complex narrative task. The participants were grouped into high and low working memory groups and their performance on the narrative tasks was correlated with the scores of working memory they obtained. The first narrative task involved the description of a comic strip consisting of six pictures. The pictures were presented in the correct order and the story line was given. In the second narrative task participants were asked to tell a story based on six unrelated pictures, all of which had to be included in the narrative. The second narrative task was obviously more cognitively demanding than the first narrative task taking into account that learners had to rely on their language skills and use their imagination to invent a story to relate the pictures to one another. Trebits and Kormos' (2008) results revealed that participants in the high complex WM group performed better on the most complex task (i.e. the second narrative), with a correlation being found between BDS test scores and both fluency and accuracy on the complex version of the narrative task. In terms of accuracy and syntactic complexity on the two tasks, no differences were found between participants with low and high working memory capacity. The interpretation of these results according to the researchers is that learners with high working memory capacity may have larger L2 vocabulary and can regulate their attention more efficiently than those with low working memory capacity in more cognitively complex tasks.

With a similar research design as that of Trebits and Kormos (2008), Gilabert and Muñoz (2010) employed an L1 reading span task developed for both Spanish and Catalan, a film retelling task, and the standardized Oxford Placement Test. Fifty-nine undergraduate university students with L1 Catalan/Spanish with high-intermediate/advanced proficiency levels of English were recruited in the study. For the film retelling task, a clip of Chaplin's *Modern Times* was used to elicit participants' L2 speech. As for learners' L2 performance on this narrative task, four different measures of performance were used each tapping into a different dimension: fluency, lexical complexity, structural complexity, and accuracy. The results show that there is a significant correlation between complex WM capacity and overall performance on fluency, and between complex WM and overall performance on lexical variety, but no significant correlations were found between working memory and structural complexity or accuracy. When the group was split into low and high proficiency groups, moderate correlations were found between lexical complexity and working memory only for the high-proficiency group. As for the first two correlations between complex WM capacity and fluency and lexical complexity, the authors of the study argue that "it is possible that higher working memory capacity is associated with faster lexical access and retrieval, which as a consequence have a positive effect on fluency". They also argue that the non-significant correlations between complex WM capacity and structural complexity can be explained in the sense that it is not always the case that more proficient learners may use more complex structures.

In one more study, Kormos and Sáfár (2008) looked at the relationship between PSTM capacity and complex WM capacity and performance in the Cambridge First Certificate language exam as the proficiency measure (i.e. in the development of reading, writing, listening, and speaking) in 121 learners (100 learners were at the beginner level and 21 at the pre-intermediate level). The research was conducted in two consecutive years in a Hungarian-English bilingual secondary school in Budapest. PSTM capacity was operationalised by a non-word span test, and complex WM capacity by a BDS test. The former test was administered to all learners at the end of the academic year in both years, and it correlated with overall proficiency, writing and use of English in intermediate students, but no significant correlations were found in beginner students. As for the BDS test, only 45 beginning students took this test. The BDS test correlated very highly with overall English language competence, as well as with reading, listening, speaking, and use of English test scores.

In Linck and Weiss' (2011) study, twenty-four L1 English learners on L2 courses (eight German, sixteen Spanish) who were enrolled in their first, second, or third semester completed an operation span task and two L2 proficiency measures: The L2 Spanish learners completed twenty items from the grammar and vocabulary section of the *Diplomas de Español como Lengua Extranjera* published by the *Instituto Cervantes,* while the L2 German learners completed fifteen fill-in-the-blank items taken from the University of Wisconsin language course placement exam. For both languages participants completed separate, parallel forms at the test and retest sessions. A hierarchical multiple regression analysis was conducted on the L2 proficiency scores and complex WM. The results showed that complex WM was positively related to proficiency at the first testing session (r = .31, p = .052). The analysis of the second session (retest) data provided further evidence of the role of complex WM in L2 learning. The results showed that WM was significantly related to proficiency (r = .56, p = .001).

In a longitudinal study (Serafini & Sanz, 2015) that investigated whether the role of complex WM capacity varies over the course of L2 morphosyntactic development, 23 beginning, 33 intermediate, and 31 advanced university L2 Spanish learners completed a digit span task, a modified Operation span task (Ospan task) (Turner & Engle, 1989; Unsworth et al., 2005), and two assessment tasks adapted from Bowles (2011): (a) an elicited oral imitation task (EI) designed to measure implicit linguistic knowledge and (b) an untimed grammaticality judgement task (untimed GJT) designed to measure explicit knowledge of ten grammar points in Spanish at three points during and after a semester of instruction, that is, around the beginning of instruction, 10 weeks later at the end of instruction, and four weeks after a period without instruction. The study yielded the following results:

- a significant correlation between complex working memory and ungrammatical items on the EI test in the beginner group at the third testing time;
- a significant correlation between complex working memory and grammatical items on the EI test in the intermediate group at the third testing time;
- no significant correlations between complex working memory and grammatical and ungrammatical items on the EI test in the advanced group;
- no significant correlations between complex working memory and grammatical and ungrammatical items on the untimed GJT in the advanced group.

Serafini and Sanz (2015) interpret these results as an indication that a relationship between variation in grammatical performance and variation in cognitive ability would result for lower level learners with less exposure to and practice using the target language. They also argue that learners at higher levels may have found the task less challenging than lower level learners; this can explain the non-significant correlations. The results also reveal that a prolonged exposure to the target language may have neutralized the facilitative role of working memory capacity, and ungrammatical items are more likely to tap learners' explicit knowledge while grammatical items are more likely to tap implicit knowledge. These findings demonstrate that "the facilitative effects of cognitive ability appear to lessen with increasing L2 proficiency and empirically support a developmental perspective of L2 learning" (Serafini & Sanz, 2015, p. 1).

Sanz, Lin, Lado, Stafford, and Bowden (2014) conducted two studies (N = 23, N = 21) investigating the role of working memory capacity and the effects of learning conditions that combine input-based practice with explicit feedback but differ in the presence or absence of a pre-practice grammar lesson on the *ab initio* language development of Latin morphosyntax, which is operationalised as increased ability to correctly assign semantic functions to noun phrases in Latin, involving the processing of case and agreement morphology. In Experiment 1, the participants completed a computer-administered treatment that lasted approximately 70 minutes and consisted of (a) vocabulary presentation and testing, (b) a grammar lesson, and (c) task-essential practice with explicit feedback. The grammar lesson combined explicit rule explanations with examples; in the testing phase, learners had to choose from two choices the right translation of a sentence in Latin. L1 English young adults who were all beginner L2 Spanish learners completed an L1 listening span test adapted from Waters and Caplan's (1996) reading version of the original Daneman and Carpenter (1980) test, a Latin written test, a Latin aural interpretation test, and a Latin grammaticality judgement test on a number of Latin sentences targeting semantic functions to noun phrases. In Experiment 2, the treatment is similar to Experiment 1, minus the grammar lesson. In Experiment 1, correlational analyses between language test gain scores and WM scores did not yield any statistically significant correlation. In contrast to this result, in Experiment 2 (with no pre-practice grammar lesson) the results show significant correlations between WM scores and pre-post gain scores from the aural interpretation test and the pre-delayed gain scores from the written interpretation test. It appears then that "when providing learners with a pre-practice grammar lesson, WM capacity does not predict language development" (Sanz et al., 2014, p. 18). It may be possible that when learners receive grammar explanations of grammar points, their processing capacity is not challenged, and therefore IDs in WM are irrelevant.

In one more study, Roehr and Gánem-Gutiérrez (2009) measured learners' explicit knowledge on a number of grammar points. This study included a total of 39 L1 English university-level L2 learners in a British university (19 students of L2 German and 20 students of L2 Spanish). The learners completed a test of metalinguistic knowledge in German or Spanish designed by the researchers consisting of two sections: the first section measured learners' ability to explicitly describe and explain aspects of the L2, operationalised as the ability to correct, describe, and explain highlighted sentence-level errors, and the second section tested learners' L2 language-analytic ability, operationalised as the ability to identify the grammatical role of parts of speech in L2 sentences. Learners were also asked to complete the MLAT (Carroll & Sapon, 2002), and a reading-span test in L1 English (Daneman & Carpenter, 1980). Correlations (Pearson's *r*) were calculated between participants' performance on the MLK test and the reading-span test, and between language aptitude and working memory. In addition to this correlational analysis, a principal components analysis was conducted on the subcomponents of the MLK test, the two measures of working memory, and the sub-tests of the MLAT. With regard to the correlational results, the L1 reading span did not correlate significantly with any of the metalinguistic measures, and a significant correlation was found between language learning aptitude as measured by the overall MLAT performance and L1 reading span (r = 0.40). On the other hand, with respect to the principal components analysis, the subcomponents of the metalinguistic knowledge loaded on component 1, the working memory measure loaded on component 3, and the subcomponents of aptitude loaded separately on components 2 and 4. Roehr and Gánem-Gutiérrez argue that a possible reason for the lack of a significant relationship between metalinguistic knowledge and working memory for language is the type of measurement employed. In other words, the MLK test was not timed and the WM test was timed having learners perform under time pressure in one measure but not on the other. It is important to mention that the L2 proficiency of learners was not specified and that beginners were not included in this study. Concerning the principal component analysis, the individual differences (language aptitude and working memory) loaded separately on distinct components, indicating (as argued by the researchers) that they each constitute a separate construct.

2.4.5 The relationship between WM capacity and language aptitude

Working memory has been discussed as a potential component of language aptitude in SLA (Miyake & Friedman, 1998; Robinson, 2005; Skehan, 2002), but few studies have addressed this issue empirically (Robinson, 2002; Roehr & Gánem-Gutiérrez, 2009; Sáfár & Kormos, 2008; Yalçin, Çeçen, & Erçetin, 2016).

Robinson (2002) studied Reber's (1993) claim that implicit learning is insensitive to IDs in cognitive abilities, while at the same time, Robinson examined the generalisability of Reber's claims for implicit learning to adult incidental L2 learning. Robinson used an experimental, repeated measures design in which 55 participants, all undergraduate third and fourth year students at a Japanese university, were recruited. There were 38 participants in the experimental group and 17 in the control group. The experimental group completed three experimental tasks: an explicit learning task (forced choice, series-solution problem task), an implicit learning task (memory task – strings of letters generated by a miniature artificial "Reber" grammar), and an incidental learning task (vocabulary learning task). On the first task, learners had to complete incomplete series of letter strings. On the second task, learners viewed strings of letters and were instructed to write them down on a piece of paper. On the last task, learners viewed sentences of the ergative, locative and incorporated type. Three testing times were planned during the study in which participants completed grammaticality judgement tests: (a) immediate post-test, (b) one week delayed post-test, and (c) six month delayed post-test. They also completed three individual differences measures: an intelligence test (a short form of the Wechsler Adult Intelligence Scale (WAIS-R)), Sasaki's (1996) Language Aptitude Battery for the Japanese (LABJ), and Osaka and Osaka's (1992) reading span test. The control group was used to calculate effect size for incidental learning; this group only completed the vocabulary learning phase and the immediate post-tests. The results with respect to the measures of WM capacity and language aptitude revealed a moderate correlation ($r = .35$, $p < .05$). With respect to the variables WM and implicit and explicit knowledge, no significant correlations were found; in contrast, a significant correlation was found between language aptitude and explicit knowledge.

In another experimental study (Sáfár & Kormos, 2008), 40 students of an English-Hungarian bilingual secondary school (experimental group) and 21 students in a regular Hungarian secondary school (control group) were recruited. The experimental group had sixteen 45-minute English lessons per week, and 4 additional English for special purposes lessons. The control group did not receive special instruction in foreign languages in their school, and they had four 45-minute lessons per week in a foreign language. The teaching method used in both groups was predominantly communicative combined with focus-on-form instruction. Language aptitude as measured by the Standard Hungarian Language Aptitude Test (HUNLAT- a Hungarian version of the MLAT) and phonological short-term memory as measured by a non-word span test were assessed both at the beginning and the end of the academic year in both schools. A BDS test was only completed by the students in the bilingual school. In the case of the experimental group, the researchers found no significant correlation between language aptitude and learners' performance on the non-word repetition task but did find a significant moderate correlation between language aptitude and learners' performance on the BDS test in the first ($r = .36$, $p < .05$) and second administration ($r = .34$; $p < .05$) of both measures. There was also a significant moderate correlation between the BDS test and the HUNLAT subcomponent language analysis ($r = .33$, $p < .05$) in the first administration of both measures. In the case of the control group, no correlational analysis was carried out because the number of students in this group was not sufficient.

In Yalçin, Çeçen, and Erçetin's (2016) study, 72 Turkish university students with advanced English proficiency (with a minimum total score on the paper-based TOEFL of 550 points) were recruited. Moreover, they had successfully graduated from a teacher training high school and were enrolled in an undergraduate degree program in English language teaching and were preparing to become English language teachers. Three instruments were used to measure complex WM capacity: a reading span task in L1 Turkish, a reading span task in L2 English, and an operation span task in L1. Language aptitude was measured with the LLAMA test (Meara, 2005). A correlational analysis indicated that complex WM capacity significantly correlated with the overall language aptitude score (reading span task in L1, r = .24; reading span task in L2, r = .30) and with the LLAMA_F subtest (r = .41 and r = .32, respectively). Although complex WM capacity and the LLAMA_F subtest were significantly correlated, the working memory tasks loaded together on a single component, separate from the aptitude subtests. This finding does not endorse the hypothesis of working memory as a component of language aptitude. This is consistent with Roehr and Gánem-Gutiérrez'(2009) findings.

Whether working memory capacity is a component of language aptitude (Robinson, 2005; Sanz, 2005; Sawyer & Ranta, 2001) as suggested by the results of some studies (Robinson, 2002; Sáfár & Kormos, 2008) is still open to question. Other studies suggest the opposite (Granena, 2013; Hummel, 2009; Roehr & Gánem-Gutiérrez, 2009b; Yalçin, et al., 2016), that is, working memory and language aptitude are different constructs. It can be argued that learners possess very different skills for WM capacity and for language aptitude (Hummel, 2009), working memory cannot be equated with language aptitude, and it can thus be said that these two cognitive abilities are not interchangeable (Yalçin, et al., 2016). Yet it can also be argued that the

LLAMA_D subtest can be assumed to reflect the efficiency of creating phonological representations in working memory, and the LLAMA_F subtest involves processing of verbal material and thus the manipulation of information in working memory (see Section 2.3.3.1). Recall that working memory capacity involves the ability to manipulate and store information simultaneously, and language aptitude involves the abilities of phonetic coding, language analysis, and memory. Thus, while the two constructs are not interchangeable, it is nonetheless plausible to argue that WM capacity is implicated to some extent in certain aptitude subtests.

2.4.6 Interim summary

As discussed at the outset of this section, the PSTM and complex WM have differential roles in L2 learning. The former is related to lexical development while the latter is related to language skills such as sentence processing, reading, speaking, and general proficiency (Juffs & Harrington, 2011). The studies discussed in this section share the common denominator that they all include a complex working memory task, and some an L2 proficiency measure (Harrington & Sawyer, 1992; Kormos & Sáfár, 2008; Linck & Weiss, 2011; Mizera, 2006; Gilabert & Muñóz, 2010), one study includes measures of implicit and explicit knowledge (Serafini & Sanz, 2015), and two studies include measures of implicit and explicit knowledge of a number of grammar points and a measure of language aptitude (Roehr & Gánem-Gutiérrez, 2009; Serafini & Sanz, 2015). The studies using a proficiency measure differ on the type of measure used, but nonetheless, significant correlations were found between language proficiency and working memory. On the other hand, on the studies including different groups of proficiency and measures of implicit and explicit knowledge (Gilabert & Muñoz, 2010; Serafini & Sanz, 2015; Trebits & Kormos, 2008), significant correlations were found between implicit knowledge and working memory but not between explicit knowledge and working memory. In Serafini and Sanz' (2015) study, no significant correlation was found between implicit knowledge and working memory for the advanced group. An interesting result is that of Sanz et al.'s (2014) study, in which a significant correlation was found between explicit knowledge and working memory only for the group of participants who did not receive explicit instruction.

What all these results show is that working memory plays an important role in two different dimensions: (a) high WM learners perform better on complex tasks, and (b) WM relates differently to L2 outcomes as proficiency increases. An overall possible explanation for the variation in results may be the type of WM tests and proficiency measures used in each study; that is, both types of measures may trigger in learners similar or different processing mechanisms given the results in the studies discussed in this section. The first dimension (a) suggests that high WM learners are better performers on complex tasks may be due to larger L2 vocabulary and more efficient regulation of attention on the part of the learners (Trebits & Kormos, 2008), and it is also possible that high WM learners may have a larger repertoire of grammar rules.

The second dimension (b) implies that the variation in results for the association between IDs in working memory and performance on implicit and explicit measures is possibly caused by the quantity and quality of L2 speaking practice that learners engage in every day, and that the development of speaking skills may tax other faculties more than working memory (Mizera, 2006). The type and length of instruction may play a crucial part in this association, that is, explicit instruction, for instance, may neutralize the role of complex WM capacity (Serafini & Sanz, 2015) though the evidence that high-proficient learners may not benefit from working memory seems inconclusive given the variation of findings.

2.5 Overall summary

As discussed in this literature review, a number of studies have been conducted on the variables language proficiency, implicit and explicit knowledge, learning difficulty, and individual differences in working memory and language aptitude. A number of researchers (Absi, 2014; R. Ellis, 2006; Robinson, 1996; Rodríguez Silva & Roehr-Brackin, 2016; Roehr & Gánem-

Gutiérrez, 2009a; Yalçin & Spada, 2016) have used measures of implicit and explicit knowledge containing grammar points that vary in learning difficulty. The grammar points included on the outcome measures have been categorized following different criteria. For instance, Yalçin and Spada (2016) used the criteria of formal complexity, input frequency, salience differences and learners' judgements, whereas Robinson (1996) used only the teachers' judgements. R. Ellis (2006) used the criteria of frequency, saliency, functional value, regularity and processability as implicit knowledge, and conceptual clarity and metalanguage as explicit knowledge. Roehr and Gánem-Gutiérrez, (2009a) drawing on R. Ellis (2006) and DeKeyser (2005) developed a taxonomy for assessing implicit and explicit learning difficulty and used it in their study. What all these criteria inform is a no consensus in the research literature with regard to what makes the acquisition of different grammar points more or less difficult, particularly when the results are inconclusive in relation to implicit and explicit knowledge of grammar points.

Two of the studies listed above (Absi, 2014; Rodríguez Silva & Roehr-Brackin, 2016) employed learner and teacher judgements to classify a number of grammar points. On both studies learners scored better on easy grammar points than on difficult grammar points. It appears that a categorization of grammar points based on learner and teacher judgments provides a fuller representation of the difficulty of grammar points. The current study aims to obtain both teachers' and learners' perceived difficulty judgements on a range of grammar points and investigate the relationship between these judgements and learners' performance on measures of implicit and explicit knowledge. Unlike the studies cited above, it examines learners' performance on three levels of proficiency (intermediate, Level 5; upper-intermediate, Level 7; and advanced, Level 9).

With regard to the relationship between explicit and implicit knowledge, the results of different empirical studies (Fotos, 1993; Akakura, 2012; Alipour, 2014; Absi, 2014; Cerezo et al., 2016;

Roehr, 2008) indicate that explicit knowledge correlates with implicit knowledge to some extent, and the results of analysis by grammar point show that both types of knowledge do not develop concurrently. It has been discussed in the review of the literature that the use of implicit measures has increased in empirical studies in recent years (Goo et al., 2015; Spada & Tomita, 2010) and the use of explicit measures has decreased. Among the implicit measures, researchers have used EI tests, timed grammaticality judgement tests, and oral narrative tests; among the explicit measures, researchers have used the untimed grammaticality judgement tests and the MLK tests. The validation studies have proved that these implicit and explicit measures are valid and reliable (R. Ellis, 2006; Erlam, 2006; Gutiérrez, 2013).

The present study examines the relationship between explicit and implicit knowledge of 13 selected grammar points in an under-researched context with L1-Spanish intermediate, upper-intermediate, and advanced L2-proficient learners of English and uses an elicited imitation test and an oral narrative test to measure implicit knowledge and a metalinguistic knowledge test to measure explicit knowledge. The current study combines the EI test and ON test scores to obtain a fuller representation of learners' implicit knowledge. With respect to explicit knowledge, a MLK test was used instead of a timed/untimed grammaticality judgement test because according to some findings (R. Ellis, 2006; Gutiérrez, 2013) it is still not 100% clear what exactly the items on these tests measure.

Research on language aptitude and working memory capacity pertaining to the acquisition of easy and difficult grammar points, on the other hand, is still relatively sparse, and the relationship between aptitude and implicit and explicit knowledge of easy and difficult grammar points, conflicting results have been found (Robinson, 1997). The different findings in the studies reviewed in the literature review suggests that different levels of L2 acquisition correlated with language aptitude (Sáfár & Kormos, 2008; Kormos & Sáfár, 2008; Roehr &

Gánem-Gutiérrez, 2009b) and that high levels of L2 acquisition correlated with high language aptitude (Forsberg & Sandgren, 2013). Likewise, the relationship between working memory and implicit and explicit knowledge of easy and difficult grammar points is relatively sparse. Results show significant correlations between implicit knowledge and working memory but not between explicit knowledge and working memory for lower levels of proficiency (Gilabert & Muñoz, 2010; Serafini & Sanz, 2015; Trebits & Kormos, 2008). To my knowledge, only two studies include the variables grammar points, language aptitude, and working memory (Erlam, 2005; Roehr & Gánem-Gutiérrez, 2009b). In the former, only one grammar point was included (direct object pronouns in French), and in the latter a number of L2 German and L2 Spanish grammar points were included but they were not categorized into easy or difficult grammar points. The current study is a further attempt to investigate the relationship between language aptitude and working memory and implicit and explicit knowledge of easy and difficult grammar points at three different levels of L2 proficiency, particularly, two hypotheses were formulated on the possible effect language aptitude and/or working memory may have on the explicit and implicit knowledge of difficult and/or easy grammar points.

There is as yet no published research that has examined the relationship between a combined judgement of teachers and learners on learning difficulty as implicit and explicit knowledge of targeted grammar points, learners' performance on measures of both explicit and implicit knowledge, and learners' performance on measures of language learning aptitude and working memory capacity in three different levels of L2 proficiency (intermediate, upper intermediate, and advanced). The present study is aimed at addressing this gap in the literature. In addition, this study addresses the possible relationship between participants' background, namely, use of English at home, use of English at work, attendance to self-access centre and implicit and explicit knowledge. The rationale for this is that it is usually the case that a number of learners seek to further develop their language skills by using the language at work or home if this is a possibility for them; it is likely that these variables could be associated to learners' outcome in L2 learning.

2.6 Research questions

The present study addresses the following research questions:

1. How did the teacher and learner participants judge the targeted grammar points in terms of learning difficulty?

2a. What is the level of learners' explicit knowledge of the targeted grammar points as measured by the metalinguistic knowledge test?

2b. What is the level of learners' explicit knowledge of easy and difficult grammar points as measured by the metalinguistic knowledge test?

3a. What is the level of learners' implicit knowledge of the targeted grammar points as measured by the oral elicited imitation and the oral narrative test?

3b. What is the level of learners' implicit knowledge of the easy and difficult grammar points as measured by the oral elicited imitation and the oral narrative test?

4. What is the relationship between learners' implicit and explicit knowledge of the 13 targeted grammar points?

5. What is the relationship between learners' implicit and explicit knowledge of the 13 grammar points, language learning aptitude and working memory capacity?

6a. Do language learning aptitude and working memory predict learners' explicit and implicit knowledge of difficult grammar points?

Hypothesis 1: Language learning aptitude and/or working memory will have an effect on the explicit and implicit knowledge of difficult grammar points, but not of easy grammar points.

6b. Do language learning aptitude and working memory predict learners' explicit and implicit knowledge of difficult grammar points at different levels of proficiency?

Hypothesis 2: Language learning aptitude and/or working memory will have an effect on the explicit and implicit knowledge of difficult and easy grammar points in the lower proficiency group, but not in the higher proficiency groups.

7. Is L2 use outside the classroom related to learners' performance on the measures of explicit and implicit knowledge of the targeted grammar points?

CHAPTER THREE: METHODOLOGY

3.1 Research design

The research design is a cross-sectional, correlational study to investigate the relationship between a combined judgement of teachers and learners on learning difficulty as implicit and explicit knowledge of targeted grammar points, learners' performance on measures of both explicit and implicit knowledge, and learners' performance on measures of language learning aptitude and working memory capacity in three different levels of L2 proficiency (intermediate, upper intermediate, and advanced). In addition, the present study addresses the possible relationship between participants' background, namely, use of English at home, use of English at work, attendance to self-access centre and implicit and explicit knowledge; three tests were developed for this purpose. In addition to these measures, IELTS speaking band descriptors (public version) were used as a determinant of learners' L2 proficiency levels, and two difficulty judgement questionnaires were employed to measure teachers' and learners' judgements of learning difficulty of the selected grammar points. Two tests were also used to further investigate learners' language learning aptitude and working memory capacity.

3.2 Participants

Two groups of participants participated in this study; a group of 101 learners and a group of 26 teachers teaching and learning English in a university in the state of Aguascalientes, Mexico.

3.2.1 Context

The learners were learning English as a foreign language in the so-called English Extension Program of a university in Mexico. In this program the main objective is to develop the four skills (i.e. speaking, listening, reading, writing) and grammar. Classes follow the presentation-controlled practice-free production approach, providing practice in all skills, with an emphasis on communicative activities that reflect real-life language use: for example, ordering food in a restaurant, discussing cultural differences between Mexico and the United States, or researching and writing a short report about a chosen topic relevant to learners' interests, such as top football teams or popular films. Both planned form-focused work in accordance with set text-book units and reactive focus-on-form activities are followed in class. Focus-on-form activities include the explicit presentation and discussion of pedagogical grammar rules followed by controlled exercises applying the rules. Classes are conducted mainly in English, but learners' L1 (Spanish) may be used in the context of form-focused activities in particular.

The program offers nine different levels of 80 hours each lasting the whole semester (there are two semesters per year of five months each). Eight of these levels are general English courses, and the first level is a beginner course, the second an elementary course, the third and fourth pre-intermediate course, the fifth and sixth intermediate courses, the seventh and eighth upper-intermediate courses, and the last level is a TOEFL course (advanced course). The only two requirements to take classes in this program are: 1) a minimum age of 16; and 2) to take a department-internal English placement exam two weeks before the classes begin. English classes in this program start one week after all undergraduate students attend their classes in their respective areas. Taking into account that the first requirement sets a minimum age of sixteen, the courses are intended for young adults and adults. Typically, the people taking these courses are people from the community, university students and university staff (i.e. both teachers and administrative personnel). This, in turn, leads to potentially mixed groups, that is, people from different backgrounds, language learning experience, age, and so on.

Once a learner has been placed in a specific course, they have to pass achievement exams and get scores equal to 7 or higher (in a scale from 0 to 10 points) and do activities at a self-access centre for a total of 8 hours in order to pass the course. The eight hours at the self-access centre represent 10 percent (1 point) of the final grade. To do the relevant activities at the self-access center, learners have to activate an identification card at the Registry Department of the university. With this card, every time they attend the self-access centre, they have to check in and out. Learners are only allowed to attend the self-access centre for a maximum of two hours per week. The activities they can do include doing exercises related to grammar or any of the basic language skills, reading hours, watching videos, listening to dialogues by using headsets, researching linguistics topics in internet, getting some language support by tutors, and attending conversation clubs which are organized by course level, that is, there are conversation clubs of one hour each at the same hour at different times of the day for beginner, elementary, pre-intermediate, intermediate, upper-intermediate, and advanced courses. For every area they wish to work on they get a handout (e.g. a listening or reading handout) with exercises on some specific topic which they have to complete, and a teacher tutor signs the completed forms for the learners to hand them to the class teacher. Every handout includes the area, language, level of proficiency of the activity, topic, objective, and possible time required to complete the activity. This procedure does not apply for conversation clubs for which learners have to register in advance at the end of each conversation club, the teacher in charge gives learners a signed notification sheet for the class teacher to confirm the learner’s attendance to that particular conversation club. Learners who do not achieve a final score of 7 in the course, or those who do achieve a score of 7 but do not do the 8 compulsory hours at the self-access centre fail the course and have to repeat the same course. Those learners who obtain a score of 7 or higher and completed the 8 hours at the self-access centre can proceed into the following level/course.

The department-internal English placement exam comprises a written multiple choice test on grammatical aspects. A short interview only takes place if a test taker’s score is on the boundary line of number of points (one or two points off) needed to be placed on a higher level in the English Extension Program. The test comprises from 10 to 15 items for each level and lasts from

10 to 15 minutes. There is no reading, writing, nor listening component. This placement test places learners in different levels/courses in the English Extension Program.

Some of the participants might have taken English classes since primary school and other participants might have studied some courses in private English schools. In other cases the participants might have taken the four compulsory English courses for university students as part of their curricula. One major issue that all these types of learners face is that there are few opportunities to practice their listening and speaking skills with native or native-like speakers once they are outside of the classroom. Nevertheless, a number of learners might be highly proficient speakers of English because they either were born or lived in the United States or Canada for many years and might use the English language at home. It is also usually the case that some learners who enrol in the English courses study the degree of Bachelor's of Arts in English Language Teaching (B.A. in E.L.T.). These learners might be working as teachers in public or private institutions. This means they use the English language at work.

3.2.2 Learner participants

The group of participants comprised 101 learners in the English Extension Program in three different groups: one group at Level 5 (intermediate, $n = 35$), one group at Level 7 (upper-intermediate, $n = 35$), and one group at Level 9 (advanced, $n = 31$); 30 students of each level completed all tests. Two of the groups, Level 5 and 7 were studying English from Monday to Friday (fifty minutes per day), and two more groups of the same level were studying English on Saturdays (five hours every Saturday). The two groups of Level 9 were studying English on Saturdays. The information that the learners provided on the background information questionnaire (see Appendix A) is reported in what follows. Table 3.1 shows the descriptive statistics for the variable age of the sample of learners in the three groups.

Table 3.1 Descriptive statistics for the learners

	Level 5	**Level 7**	**Level 9**
Number of participants	30 (21 female and 9 male students)	30 (18 female and 12 male students)	30 (21 female and 9 male students)
Age			
Range Mean Median	16-67 34.9 30	17-54 26.4 20.5	17-69 26.2 23.5
Years of formal L2 study			
Range Mean Median	1-12 2.9 2	2-18 5.7 4	1-17 7.4 5

Two learners in Level 5 reported they had lived in Canada for one year, and four learners of Level 9 reported they had spent eight months, two, seven and eleven years respectively living in the United States. Two more learners in Level 9 reported they had lived in the same country for ten years. These learners' scores were checked on both implicit and explicit measures and their language experience did not show extreme scores that could bias the distribution of the data.

The learners also reported studying or having studied one of the following university degrees in the same university.

Table 3.2 Frequencies for number of learners studying a university degree

University degree	**Level 5**	**Level 7**	**Level 9**
ELT	0	0	8
Accounting	0	3	2
Medicine	2	4	1
Biology	1	0	0
Law	1	0	0
Industrial Engineering	1	0	0
Computing	1	0	2
Architecture	0	1	1

Automotive Engineering	0	2	1
Education	0	0	1
Psychology	0	1	0
Biochemistry	0	1	0
Civil Engineering	0	1	0
Electronic Engineering	0	1	0
Master's in Nutrition	1	1	0
High school	5	3	2

From all these university degrees, only one learner reported his intention to start the BA in ELT at the Language Department of the university in 12 months. Furthermore, eight learners reported currently studying in the BA in ELT in the same department: one learner reported being enrolled in the first semester, one in fifth, and six in seventh.

With regards to other languages learners had learned, Table 3.3 displays seven different languages for all groups.

Table 3.3 Frequencies for number of learners studying other languages

	Level 5	**Level 7**	**Level 9**
French	4	5	2
Italian	1	4	2
German	0	4	0
Japanese	0	3	4
Portuguese	0	1	0
Chinese	0	1	0
Korean	0	0	1
French and Italian	0	2	0
French and German	0	2	0
French and Chinese	0	1	0
Japanese and Korean	0	0	1

Table 3.3 shows that the total number of learners in Level 7 studying other languages apart from English as an L2 was higher than the other two groups. Two learners in Level 7 studied French and Italian, two more French and German, and one French and Chinese. In Level 9, only one learner studied Japanese and Korean. The performance of these learners on the implicit and explicit measures (i.e. MLK, EI and oral narrative test) and LLAMA and WM test did not represent outlying scores.

With respect to the use of English at home, two learners in level 5 reported to speak English with brothers and children, six learners in level 7 with parents, brothers, and children, and five learners in level 9 with parents, brothers, and cousins. Concerning the use of English at work, five learners in level 7 reported to speak English with co-workers and clients, and six learners in level 9 reported to speak English with clients, students and teachers.

Regarding the requirement of the 8 compulsory hours at the self-access centre, 29 learners reported having attended the conversation clubs at self-access centre. Twenty three learners reported 1 hour per week, four 2 hours per week, and two 2 hours per week.

3.2.2.1 L2 proficiency level

In an effort to employ robust proficiency assessment standards and to confirm the L2 proficiency level of the learners according to their level groups (Level 5, 7 and 9), inferential statistics were conducted using their L2 proficiency scores and years of formal L2 study. Table 3.4 displays mean performance scores and years of formal L2 study, standard deviations from the mean, and minimum and maximum scores.

Table 3.4 Descriptive statistics of the formal L2 study and L2 proficiency scores

Formal L2 study	N	Mean	Median	SD	Min	Max
Level 5 (Intermediate)	30	2.58	2	1.60	1	9
Level 7 (Upper-intermediate)	30	3.73	3.50	1.34	2	7
Level 9 (Advanced)	30	4.87	4	2.15	2	10
L2 proficiency						
Level 5 (Intermediate)	30	5.53	6	1.14	4	8
Level 7 (Upper-intermediate)	30	6.10	6	1.37	4	9
Level 9 (Advanced)	30	6.63	7	1.10	5	9

In order to find out whether there were any significant differences between the level groups and given that the distribution of the data for the variable years of formal L2 study was non-normal ($p < .05$), a non-parametric test (Kruskal-Wallis test) was run for each variable. The Kruskal-Wallis test showed that there was a statistically significant difference between the different level groups, X^2 (2) = 28.26, $p < .01$ for formal L2 study. A further analysis was conducted to see whether Level 5 differed from Level 7, Level 5 from Level 9, and Level 7 from Level 9. The Mann-Whitney test indicated that Level 7 was greater (*Mdn* = 38.68) than Level 5 (*Mdn* = 22.32), U = 204.50, $p < .01$; Level 9 was greater (*Mdn* = 41.20) than Level 5 (*Mdn* = 19.80), U = 129, $p < .01$, and Level 9 was greater (*Mdn* = 35.92) than Level 7 (*Mdn* = 25.08), U = 204.50, $p < .05$.

Another Kruskal-Wallis test was run for mean scores of L2 proficiency (the distribution of the data was non-normal). The Kruskal-Wallis test showed that there was a statistically significant difference between the different level groups, X^2 (2) = 11.25, $p < .01$ for L2 proficiency. A further analysis was conducted to see where the differences lay. The Mann-Whitney test indicated that Level 7 was greater (*Mdn* = 33.92) than Level 5 (*Mdn* = 27.08), U = 347.50, p = .119; Level 9 was greater (*Mdn* = 37.92) than Level 5 (*Mdn* = 23.08), U = 227.50, $p < .01$, and

Level 9 was greater (*Mdn* = 34.27) than Level 7 (*Mdn* = 26.73), U = 337, p = .085. These significant differences indicate that the each group of learner participants had a different level of English proficiency.

3.2.3 Teacher participants

The group of participants comprised 26 teachers teaching English at different levels on the English Extension Program in the same university. The background information about the teacher participants is shown in Table 3.5.

Table 3.5 Descriptive statistics for teacher participants

Number of participants	**Age**	**Language teaching experience (years)**
26 (18 female and 8 male teachers)	Range: 22 - 43 Mean: 31.7 Median: 30.5	Range: 1 - 19 Mean: 8.7 Median: 8.0

Table 3.5 shows that there was diversity in age and teaching experience among teachers.

3.3 Targeted grammar points

At the onset of the present study, 13 grammar points were selected from the English textbooks (Kay & Jones, 2003, 2009) used by the participants in preceding years. The 13 grammar points included in this study were selected because they vary in degree of learning difficulty, the participants had encountered these grammar points in their classes, and items for both the implicit and explicit measures could be constructed for each of the grammar points. The targeted grammar points are presented in Table 3.6. Each grammar point is followed by a pedagogical grammar rule, an example sentence, and a typical learner error; all grammar rules were adapted from pedagogical grammar books (Biber, Conrad, & Leech, 2002; Eastwood, 1999; Murphy, 2006; Swan, 1980) which contained metalinguistic descriptions of a range of L2 grammar points. In order to facilitate test design and subsequent scoring, the pedagogical grammar rules corresponding to the thirteen grammar points were converted into the format "When form X occurs/function X is being expressed, form Y needs to be used" (Roehr, 2008; Thepseenu & Roehr, 2013; Ziętek & Roehr, 2011). These pedagogical grammar rules were modelled on the instrument used by Ziętek & Roehr (2011).

Table 3.6 Targeted grammar points selected from participants' textbooks

Grammar point	**Description**	**Example sentence(s) (focus of the grammar point is in bold)**	**Common error (underlined) made by Mexican learners**
Simple past tense (*-ed* form)	When a finished action or event in the past is being expressed, the simple past tense is required.	He visit**ed** his brother yesterday.	*When he finished his homework he <u>watch</u> a movie.
2nd conditional (*if*-clause)	When an unreal/hypothetical situation is being expressed, the 2nd conditional comprising an *if*-clause with a past tense verb and a main clause with *would* + infinitive is used.	If I **had** money, I **would buy** a car.	*If I <u>know</u> the answer, I would tell you.
3rd person –*s* in the simple present tense	When a verb in the 3rd singular person is used in the simple present tense, an –*s* or –*es* is added to the end of the verb of the sentence.	Alex want**s** to go home.	*Sara <u>cook</u> every day.
Comparative adjectives	When making a comparison, you either add *-er* to a one-syllable adjective or you place *more* in front of an adjective with two or more syllables.	Carlos is tall**er** than his sister. My book is **more** expensive than yours.	*My car is <u>more cheap</u> than yours.
Infinitives and gerunds (as verb complements)	When the main verb of a sentence is, e.g. *decide*, *hope*, or *plan*, and when it is followed by another verb, the *to*-infinitive construction is required for the second verb, but if the main verb is *enjoy*, *avoid*, or *deny*[3], the *ing*-form construction is required for the second verb.	He decided **to write** a story. She enjoys **driving** around the country.	*The boys want <u>buy</u> a new car. *They finished <u>to build</u> the house.
Indefinite article	When a countable noun is first mentioned, an indefinite article is required.	They had **a** good class today.	*She bought <u>the</u> new house.
Modal verbs + verb	When a modal verb such as *must*, *should*, or *can* is used, it is followed by the infinitive of the main verb.	I should **wait** for my brother.	*I must <u>to go</u> to work.
Many vs. much	When the quantity of something is being referred to, *many* is required for countable	She has **many** activities to do	*I have <u>many</u> money.

[3] The verbs *decide, hope, plan, enjoy, avoid* and *deny* are frequently used by the learners.

	nouns and *much* is required for uncountable nouns.	during the school term. They don't have **much** time.	*I didn't see much people at school today.
Plural of nouns	When the plural of a regular noun is being expressed, an *–s* needs to be added to the noun.	It takes a few minute**s** to get to the airport.	*The exam will start in five minute.
Yes/no questions	When a yes/no question with the auxiliary verb "do" is used, the infinitive of the main verb is required.	**Does** Maria **like** the new house? **Did** he **go** to the park?	*Does Pedro works late? *Did they took the book?
Since/For	When the specific time of the beginning of an action is expressed, *since* is required, but when the length of time of an action is expressed, *for* is required.	Jane has been in hospital **since** Tuesday. People have used mobile phones **for** many years.	*I have been here for 9 o'clock this morning. *Teachers have used computers since two decades.
Direct and indirect objects (Dative alternation)	When an indirect object follows a direct object in a sentence, the preposition *to* is placed in front of the indirect object.	The man gave a letter **to** the boy.	*The postman gave the letter the woman.
Relative clauses	When a relative clause where the relative pronoun functions as an object is used, a pronoun that makes reference to the subject of the sentence (resumptive pronoun) is not permitted.	**The table** that I saw the other day **is** expensive.	*The car that my father bought it is new.

The 13 grammar points in Table 3.6 were assessed by means of two tests of implicit knowledge and one test of explicit knowledge. These measures are described in the following section.

3.4 Instrumentation

The three different instruments developed to measure implicit and explicit knowledge were two implicit knowledge tests, and an explicit knowledge test. One implicit knowledge test was designed in an elicited imitation format (EI test) and the other one in an oral narrative format (ON test), whereas the explicit knowledge test was a metalinguistic knowledge test (MLK test). In order to determine the proficiency level of each learner, 6 teachers (including myself) employed the IELTS speaking band descriptors (public version) (www.ielts.org/microteaching/assets/docs/Speaking%20Band%20Descriptors%20V2.pdf) to assess the short story produced by each learner in the oral narrative task. Three criteria were used for this purpose: (a) fluency and coherence, (b) grammatical range and accuracy and, (c) pronunciation. The two instruments employed to measure language learning aptitude and working memory capacity were the LLAMA test (Meara, 2005), and the backward digit span test (Kormos & Sáfár, 2008). Furthermore, two difficulty judgement questionnaires were designed to measure teachers' and students' perceived level of difficulty of the selected grammar points.

3.4.1 Difficulty judgement questionnaire for teachers

In the difficulty judgement questionnaire, teacher participants were asked to give their opinion of each item according to their experience in learning and teaching the language. The questionnaire followed the same format as shown in Table 3.6 with one more added column containing a 5-point scale (very easy – easy – moderate – difficult – very difficult) (DeKeyser, 2003) for teacher participants to judge the level of learning difficulty of each grammar point holistically. An example is provided below (see Appendix B for the full questionnaire).

Example:

Grammar Point	**Pedagogical grammar rule**	**Example sentence(s) (targeted form is in bold)**	**Typical learner error (error is underlined)**	**Level of difficulty**				
				Very easy	Easy	Moderate	Difficult	Very difficult
Simple past tense (*-ed* form)	When a finished action or event in the past is being expressed, the simple past tense is required.	He visit**ed** his brother yesterday.	*When he finished his homework, he watch a movie.					

3.4.2 Measures for learner participants

The 13 grammar points in Table 3.6 were assessed by means of two tests of implicit knowledge (i.e. an EI test and an ON test), one test of explicit knowledge (i.e. a MLK test), and a difficulty judgement questionnaire. Student participants also completed the LLAMA test, and the backward digit span test.

3.4.2.1 Difficulty judgement questionnaire for learners

Like the difficulty judgement questionnaire for teachers, student participants were asked to give their opinion of each item according to their experience in learning the language. The questionnaire had the same format and content as the one completed by the teachers (see Appendix C for the full questionnaire).

3.4.2.2 Elicited imitation and oral narrative test

The construct of implicit knowledge was operationalised following R. Ellis' (2005) seven criteria. The criteria and their operationalisations in terms of implicit and explicit linguistic knowledge are presented in Table 3.7 (repeated here for convenience).

Table 3.7 Operationalising the constructs of L2 implicit and explicit knowledge

Criterion	Implicit knowledge	Explicit knowledge
Degree of awareness	Response according to feel	Response using rules
Time available	Time pressure	No time pressure
Focus of attention	Primary focus on meaning	Primary focus on form
Systematicity	Consistent responses	Variable responses
Certainty	High degree of certainty in responses	Low degree of certainty in in responses
Metalinguistic knowledge	Metalinguistic knowledge not required	Metalinguistic knowledge encouraged
Learnability	Early learning favoured	Late, form-focused instruction favoured

The two implicit measures (i.e. the EI and ON test) employed in the present study were selected due to their compliance to the aforementioned criteria, particularly to the compliance of degree of awareness, time available, focus of attention, and metalinguistic knowledge (see R. Ellis, 2005 for an overview).

The elicited imitation test used in the study was adapted from Erlam (2006). Erlam points out that an elicited imitation test of this type would be designed to require a primary focus on meaning rather than on form, it would include some delay between the presentation of the stimulus and repetition of the same, there would be some spontaneous correction of ungrammatical sentences and it would be completed under time pressure. In other words, an EI test with these characteristics may ensure that implicit knowledge is measured.

Following R. Ellis et al. (2009), the EI test was described to participants as a "belief questionnaire" to maximize the possibility that learners would be attending to the meaning rather than the form of the statements they would hear. The test includes a short training section in which the test takers practice with six items before completing the test. The test proper contains 78 items which comprise three ungrammatical and three grammatical statements for each of the 13 grammar points; the statements were pseudo-randomized. Two statements for each grammar point were adapted from Erlam's (2006) study, and the other four were designed based on the participants' Mexican context on topics such as politics, music, sports, education, football, foreign languages, and science; all the statements (including the statements in the short training section) were pre-recorded by the researcher. In this test participants hear a statement and then they decide whether the statement is true, not true or whether they are not sure. They then need to repeat the statement in correct English. All instructions are written in English and Spanish to avoid any confusion in what participants have to do. The maximum possible score was 93 points (eight statements presented two test items and sentences in the 2nd conditional presented two clauses). Two examples are provided below (see Appendix D for the full EI test).

Example 1 (grammatical sentence):

1) "Life is very difficult for many old people." True Not true Not sure

[PAUSE]

Now repeat the statement.

Example 2 (ungrammatical sentence):

1) "English spoken in many different countries." True Not true Not sure

[PAUSE]

Now repeat the statement.

As evidence that participants did focus on meaning for every single statement they heard, four sentences were selected which participants were likely to consider "true", four sentences which they would be more likely to consider "not true" and four more sentences which would most likely elicit the response "not sure". Recall that learners were asked to indicate their "beliefs" choice on a test sheet before they attempted to repeat the statement; they had three seconds to respond to each statement. In other words, it can be hypothesized that by having learners perform this task under time pressure, they focused their attention on meaning instead of rehearsing the stimulus sentence (see Table 3.8).

Table 3.8 Responses that participants made to the meaning of selected statements

Participants responses	**True**	**Not true**	**Not sure**
3. People should report stolen money the police.	76		
13. It is more harder to learn Japanese than to learn English.	69		
22. Every child needs good father.	71		
40. Teachers must prepare their classes before they give a lesson.	73		
5. The film that everyone likes is Star Wars.		49	
26. Many people study at university level today.		65	
28. A good student never study before an exam.		52	
47. President Peña Nieto has been in the presidency since 1994.		61	

14. Mijares loved Lucero but he divorced her.	48
17. If Russia had more power, the United States would be worried.	46
27. Einstein failed Math when he was a student.	51
59. Chicharito has played with Manchester United for 2006.	52

The results in Table 3.8 show evidence that the participants were indeed focusing on meaning as intended in the design of the test. For the sentences that the researcher selected believing that participants would be likely to consider "true" (statements 3, 13, 22 and 40), most participants chose the "true" option. For the other two truth values, more than half of the participants chose "not true" or "not sure" for the statements the researcher believed participants would be likely to consider "not true" or "not sure".

Like the EI test, the objective of the oral narrative test (ON test) was to measure the implicit knowledge of the targeted grammar points that participants had. This test was adapted from Ellis, et al. (2009). The test is a short story containing eight of the targeted structures (simple past tense, simple present tense, comparative adjectives, verb complements, indefinite article, modal verbs, plural of nouns, and dative alternation), a subset of the 13 structures included in the EI test and the MLK test. These eight structures were chosen because they could be incorporated in a cohesive narrative. An extract of the story is provided below; the targeted structures are highlighted for the benefit of the reader only (see Appendix E for the full test).

Example (extract):
Every morning Mr. Garcia **gets up** at 6:30 am, **walks** to the store and **buys a** newspaper. He **returns** to…

Table 3.9 shows the number of (potential) obligatory occasions for each structure included in the story that learners were asked to read and repeat orally.

Table 3.9. Obligatory occasions of eight target structures

Structures	**Obligatory occasions**
Simple past tense (*-ed* form)	12
Simple present tense (3rd person *–s*)	13
Comparative adjectives	6
Verb complements (infinitives and gerunds)	9
Indefinite article	7
Modal verbs	6
Plural of nouns	11
Dative alternation	5

3.4.2.3 Metalinguistic knowledge test

Unlike the implicit measures (EI and ON test) which mainly tap learners' implicit knowledge, a MLK test mainly taps learners' explicit knowledge. Even though it is argued that a MLK test can also tap learners' implicit knowledge if it is administered under time pressure (Han & Ellis, 1998), it can be argued that a MLK test can mainly tap explicit knowledge because "explicit knowledge is generally accessible through controlled processing" (R. Ellis, 2004, p. 237), which is usually manifested in learners describing rules being violated (Gutiérrez, 2013) and illustrating rules with example sentences.

The metalinguistic knowledge test was composed of two parts that comprised 35 items in total. The first part of the test was modelled on the instrument used by Ziętek & Roehr (2011) consisting of 18 ungrammatical sentences which covered the 13 targeted grammar points (five grammar points were represented by two test items because four grammar points involved a two-part pedagogical rule and one required two sentences as can be seen in Table 3.6: comparative adjectives, verb complements, many vs. much, since/for, and yes/no questions) for which participants had to provide the correction of a highlighted error and also provide an explanation of that correction. An example is provided below (see Appendix F for the full MLK test).

Example:

When he finished his homework, he watch a movie.

Correction: watched

Explanation: When a finished action or event in the past is being expressed, the simple past tense is required.

The second part of the MLK test presented 13 items. In this part, the targeted grammar points and their corresponding pedagogical grammar rules were provided in English and Spanish. Participants were asked to write a sentence which included the targeted L2 construction and illustrated the pedagogical grammar rule (Absi, 2014; Scheffler, 2011). An example is provided below.

Example:

Grammar point: Simple past tense

Rule: When a finished action or event in the past is being expressed, the simple past tense is required. (Please use a regular verb).

"Cuando se expresa un evento u acción terminada en el pasado, se utiliza el pasado simple. (Favor de usar un verbo regular en la oración)".

Sentence: ____

The maximum possible score for the test as a whole was 52 points.

3.4.2.4 Language aptitude test

The language aptitude test used in this study was the LLAMA language aptitude test (Meara, 2005), the most recent version of the LAT (Meara *et al*., 2001), which is designed so that it is accessible without the L1 being a factor, beyond understanding initial instructions (Skehan, 2012). Furthermore, this test is a computer-based test and it is available for free download (www.lognostics.co.uk/tools/llama).

The LLAMA test includes the following four subtests:

- LLAMA B: a test of vocabulary learning;
- LLAMA D: a test of sound recognition;
- LLAMA E: a test of sound-symbol association;
- LLAMA F: a test of grammatical inferencing.

The first subtest is the LLAMA_B test. This test is a vocabulary learning task which measures the ability to learn relatively large amounts of vocabulary in a relatively short space of time. This subtest is loosely based on the paired associates task of Carroll and Sapon (1959), but uses a different interface. This test presents a set of 20 objects on the screen representing real words taken from a Central American language which are arbitrarily assigned to the target objects.

The second subtest is the LLAMA_D test. This test is a sound recognition task which measures test takers' ability to recognize short sounds in an unknown language that they hear at the onset of the test. These sounds are generated by the computer.

The third subtest is the LLAMA_E test. It is a sound-symbol correspondence task. This test presents 24 buttons containing "the transliteration of syllables in an unfamiliar alphabet" (Meara, 2005, p. 11), and for each button that is clicked, test takers hear the syllable corresponding to that button. Test takers have to work out "the relationship between the sounds they hear and the writing system" (Meara, 2005, p. 11). They have 120 seconds "to learn how the spelling system of this language works" (Meara, 2005, p. 12), and written notes are allowed.

The last subtest is the LLAMA_F test. It is a grammatical inferencing task that presents 20 buttons and for each button that is clicked, a picture and a sentence in a constructed language that describes the picture is displayed; test takers have 300 seconds "to work out the grammatical rules that operate in the unknown language" (Meara, 2005, p. 15), and written notes are allowed.

3.4.2.5 Test of working memory

In order to assess the learners' capacity to temporarily store and manipulate information necessary for such complex tasks as comprehension, learning and reasoning (Baddeley, 2000), the backward digit span task was employed. This task, like the reading and listening span test (Daneman & Carpenter, 1980), has been used in previous studies as diagnostic tools which have been claimed to assess learners' capacity of complex verbal working memory (Kormos & Sáfár, 2008; Trebits & Kormos, 2008).

With this in mind, the backward digit span test (BDS test) was used to assess learners' working memory capacity. The test comprises seven sets of four sequences of numbers each and allows for two trials for each sequence. The first set contains four sequences of three digits such as 5, 8, 2. The following sequence in each set increases by one number, that is, the second set contains four sequences of four numbers; the following set contains four sequences of five numbers, and so on (see Appendix G for the full test).

The piloting of all the instruments took place in two different pilot studies at different times in the same learning context.

3.5 Pilot study 1

The findings reported in this section are based on the same data set underlying Rodríguez Silva and Roehr-Brackin (2016), which was reviewed in sections 2.1.6.1 and 2.2.2.3. The focus in this section is on results speaking to the quality of the instruments only.

In this first pilot study, 11 university teachers of English and 30 L1 Spanish learners of English in Level 5 (intermediate level) participated in the piloting of the difficulty judgement questionnaires, the MLK test (explicit L2 knowledge), and the oral EI test (implicit L2 knowledge). The participating teachers and learners were involved in the same English program where the main study would take place.

All participants agreed to participate in this pilot study, and they received and signed a consent form before carrying out the completion of questionnaires and tests. The learners completed the instruments under the supervision of the researcher in separate sessions in the following order: measure of implicit L2 knowledge, Part 1 of the test of explicit knowledge (correction of errors and description/explanation of grammatical rules), Part 2 of the test of explicit knowledge (production of sentences to illustrate given pedagogical grammar rules), and the difficulty judgement questionnaire. The instruments targeted 13 grammar points of English.

3.5.1 Test of implicit knowledge

The EI test included 12 out of the 13 grammar points of the last version of the test used in the main study. The grammar point preposition + verbs used in the piloting was substituted by the grammar point plural of nouns in the main study. The reasons for doing this were that preposition + verb had a floor effect on both implicit and explicit measures, and it was also more practical to use the grammar point plural of nouns on the oral narrative test which would be used in the main study as a second measure of implicit knowledge. The scoring of the EI test was as described in the main study (see section 3.7.1).

3.5.2 Test of explicit knowledge

As on the EI test, on the MLK test the grammar point preposition + verb was substituted by the grammar point plural of nouns. The MLK test was scored dichotomously, and all test items were awarded 0 or 1 point. In Part 1, 1 point was awarded for each appropriate correction. In addition, 1 point was awarded for an appropriate metalinguistic explanation reflecting the "When form *X* occurs/function *X* is being expressed" clause of the targeted pedagogical grammar rule, and 1 point was awarded for an appropriate metalinguistic description reflecting the "form *Y* needs to be used" clause of the targeted pedagogical grammar rule. In part 2, 1 point was awarded for each correctly produced sentence fully illustrating a given pedagogical grammar rule. The scoring of this test was modified in the second pilot study and subsequently in the main study.

3.5.3 Difficulty judgement questionnaire

With respect to the difficulty judgement questionnaire, the same format was used for both student and teacher participants. This instrument was modified after its administration in the first pilot study, that is, the grammar point preposition + verbs used in this pilot study was substituted by the grammar point plural of nouns. The questionnaire was administered again in the main study only; none of the two groups of participants had any problems in judging each grammar point as "easy" or "difficult" based on the 5-point scale.

3.5.4 Data analysis

Before proceeding with the statistical tests, an internal reliability analysis was conducted for the MLK test as a whole and its subsections and the EI test. The two tests showed good internal consistency (Cohen, Manion and Morrison, 2011) (EI test, Cronbach's alpha = .90; MLK test, Cronbach's alpha = .81). In addition, a comparative analysis was conducted between learner and teacher judgements on the degree of learning difficulty for each of the 13 targeted grammar points.

3.5.4.1 Difficulty judgement of grammar points

In order to address the issue on the difficulty judgement of grammar points, learner and teacher rankings of each grammar point (1 for "very easy" and 5 for "very difficult") were calculated employing the scoring scheme discussed in the scoring and coding section (see section 3.8.1). To make the interpretation of these values as to whether a construction is easy or difficult, the mean difficulty score was calculated for each grammar point. Table 3.10 shows the descriptive statistics for the learners' and teachers' judgements of the difficulty of the grammar points. The mean scores were ordered from the easiest (i.e. lowest mean scores) to the most difficult (i.e. highest mean scores) according to the learners only.

Table 3.10. Perceived learning difficulty of the targeted grammar points

Grammar point	**Learners: mean (*SD*)**	**Teachers: mean (*SD*)**
Simple present tense	1.4 (.67)	2.8 (1.4)
Simple past tense	1.5 (1.68)	2.3 (.79)
Comparative adjectives	1.6 (.67)	2.9 (1.04)
Yes/*no* questions	1.7 (.79)	2.8 (1.17)
Relative clauses	1.9 (.80)	2.5 (.93)
Many vs. *much*	1.9 (.66)	3.1 (.94)
Since vs. *for*	2.0 (.85)	2.8 (.60)
Indefinite articles	2.0 (.83)	2.8 (.87)
Modal verbs	2.2 (.73)	2.5 (.69)
Dative alternation	2.2 (1.1)	3.4 (1.12)
Second conditionals (*if* clauses)	2.5 (.78)	3.4 (.92)
Prepositions + verbs	2.5 (.97)	3.5 (.82)
Verb complements	2.9 (1.09)	3.6 (1.12)

The difficulty judgements shown in Table 3.10 indicate that the learners tended to judge the targeted grammar points to be less difficult overall than the teachers. Only one of the mean scores approaches 3 (*moderate*), whereas the majority are below 2.5 and thus in the "easy" side. By contrast, the teachers have only a single score that is below 2.5 and thus in the "easy" side. In order to ascertain whether there is any statistical relationship among the difficulty judgements made by the participant groups, a Spearman rank order correlation was run. The result confirmed that learners' and teachers' difficulty judgements are significantly correlated (rho = .63, p = .02).

3.5.4.2 Learners' performance on the measures of explicit and implicit knowledge

The results in Table 3.11 indicate that both tests were sufficiently challenging for the participants. Indeed, the EI test proved to be rather difficult, with a mean facility value of just 40%. The MLK test was somewhat easier overall. This was due primarily to the error-correction task and, to a lesser extent, to the rule-illustration task; by contrast, the description/explanation task was challenging for the learners.

Table 3.11 Descriptive statistics: Tests of implicit and explicit L2 knowledge

	EI test	MLK test	Correction	Description/ explanation	Rule illustration
Mean %	40	57	80	41	65
Mean	35.47	41.33	14.47	14.57	12.30
SD	11.43	9.10	2.43	5.82	2.82
Max. possible	89	73	18	36	19

Note. EI = elicited imitation; MLK = metalinguistic knowledge

Table 3.12 shows the correlations between scores on the two measures.

Table 3.12 Correlations (Pearson's *r*): Tests of implicit and explicit L2 knowledge

	EI test	MLK test	Correction	Description/ explanation
MLK test	.65**			
	p = .00			
Correction	.54**	.67**		
	p = .00	*p* = .00		
Description/explanation	.57**	.91**	.42*	
	p = .00	*p* = .00	*p* = .02	
Rule illustration	.46*	.76*	.45*	.52**
	p = .01	*p* = .00	*p* = .01	*p* = .00

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

As shown in Table 3.12, the MLK test correlated strongly with each of its subsections, which indicates that each of the subsections represents a coherent subordinate measure of metalinguistic knowledge. This relationship was highest between the MLK test and its subsection description/explanation where the learners had to describe or explain the error of each sentence, suggesting that this subsection mostly represented the MLK test.

Scores on the MLK test as a whole and the EI test are likewise associated, indicating that learners' implicit and explicit knowledge are related if these types of knowledge are assessed by means of measures targeting a range of L2 structures.

Table 3.13 displays the results from an analysis by targeted grammar point, thus showing the relationship between learners' performance on measures of implicit and explicit knowledge of the 13 selected L2 English grammar points, on the one hand, and the perceived learning difficulty of these grammar points as judged by teachers and learners on the other hand.

Table 3.13 Correlations (Spearman's *rho*): Perceived learning difficulty and actual performance

	EI test	**MLK test**
Learners' judgements	-.12	-.64*
	p = .71	*p* = .02
Teachers' judgements	-.53	-.52
	p = .06	*p* = .07

* Correlation is significant at the 0.05 level (2-tailed)

Although all relationships are negative, thus reflecting the expected association of higher perceived difficulty with lower scores on performance measures, there is only a single significant correlation between performance on the test of explicit knowledge and the learners' own difficulty judgements. This indicates that the learners themselves were the only participant group to make a reliable judgement, in this case for their performance on the metalinguistic knowledge test. It is worth noting that the teachers' difficulty judgements show a trend toward significance for both the test of implicit and the test of explicit L2 knowledge. Conversely, the learners are far from predicting performance on the elicited imitation test.

Finally, there is no relationship between performance on the measures of implicit and explicit knowledge in the analysis by grammar point (*rho* = .09, *p* = .76; not shown in Table 3.13), indicating that learners appear to have developed implicit knowledge of certain grammar points and explicit knowledge of others, and vice versa.

3.6 Pilot study 2

In this second pilot study, 27 students from three different level groups participated: one group of Level 5 (intermediate; *n* = 10*)*, one group of Level 7 (upper-intermediate; *n* = 7), and one group of Level 9 (advanced; *n* = 10); only 6 students of each level completed all tests. The same 13 grammar points used and modified following Pilot study 1 were assessed by means of two tests of implicit knowledge (i.e. EI test and ON test) and one test of explicit knowledge (i.e. MLK test). The EI test and the MLK test were the same measures used and amended in the first pilot study. Student participants also completed the LLAMA test, and the BDS test. As for the difficulty judgement questionnaire, I used the teachers' difficulty judgements from Pilot study 1.

The mean scores in Table 3.10 (in section 3.5.4.1) indicate that none of the grammar points was perceived as very easy or very difficult. The rating scale (1 for "very easy" and 5 for "very difficult") was reinterpreted to include two main difficulty categories: easy and difficult. The reinterpretation of the rating scale for learners was from 1.4 to 2.1 for easy grammar points and from 2.2 to 2.9 for difficult grammar points. Likewise, the reanalysis for the teachers' judgements was from 2.3 to 2.9 for easy grammar points and from 3.0 to 3.6 for difficult grammar points. Both groups of participants agreed on 11 grammar points. The grammar points many vs. much and modal verbs were perceived differently by each group of participants. Hence, the grammar point modal verbs was assigned to the easy category and the grammar point many vs. much to the difficult category by the researcher for three reasons: (1) a Spearman's rank order correlation was run between the learners' and the teachers' mean difficulty scores to find out whether there was any correlation; there was a significant and strong positive correlation between the two sets of mean scores (rho = .63, *p* = .02), (2) another study used the expert judgement of experienced L2 teachers to categorize a number of grammar points as easy or difficult (Robinson, 1996) and, (3) the results in the difficulty judgement questionnaire in pilot study 1 show that the teacher group was the only group that showed a tendency towards successful prediction of learners' performance on both explicit and implicit L2 measures. Table 3.14 shows the resulting classification of the grammar points.

Table 3.14 Difficulty of grammar points according to learner and teacher judgements

Easy	**Difficult**
1. Simple present tense	1. Many vs. much
2. Simple past tense (-*ed* form)	2. Dative alternation
3. Comparative adjectives	3. Second conditional (*if* clauses)
4. *Yes/no* questions	4. Prepositions + verbs
5. Relative clauses	5. Verb complements
6. *Since* vs. *for*	
7. Indefinite articles	
8. Modal verbs	

As in the first pilot study, all participants agreed to participate in this second pilot study, and they received and signed a consent form before carrying out the completion of questionnaires and tests. The learners completed the instruments under supervision in separate sessions in the following order: background information questionnaire, LLAMA test, WM test, EI test, ON test, and MLK test.

3.6.1 Procedure

The first measure administered was the background information questionnaire to student participants; the student participants completed the background questionnaires in fifteen minutes in one session before class was ended. Following this administration, learners completed the LLAMA test which took place in a lab in one session (approximately 60 minutes); the procedure of administration was as described in the main study (see section 3.7). This was followed by the backward digit span test on a one-to-one basis in a quiet room in one session (approximately 5 minutes with each learner). Similarly to the LLAMA test, the EI test also took place in a lab, and it was completed in two sessions during weekday classes and in one session on Saturday classes (approximately 70 minutes). The oral narrative test, as the backward digit span test, also took place on a one-to-one basis in a quiet room in four sessions during weekdays and in two sessions on Saturday classes (approximately 10 minutes with each participant). The last test administered was the MLK test. Student participants completed it in two sessions on weekdays (Part I in one session and part II in the following session; 35 minutes approximately for each session) and in one session on Saturday classes (70 minutes approximately).

For the backward digit span test, the participants first completed a training section with three and four series of digits, and then they did the test. The researcher read each series of digits to each participant in a monotone voice, one second apart. Student participants had to repeat the numbers backwards. The same series of numbers was read again, and learners had to repeat the numbers one more time in reverse order. If they repeated two series of digits correctly of a given series length, they skipped the other two series of that set and moved on to the following set of series, that is, they started with a three-digit series and moved on to a four-digit series. However, if participants failed to repeat two successive series, the test was terminated. All instructions for this test were given in Spanish (i.e., the learners' L1), and it was completed in Spanish as well.

The oral narrative test was also administered on a one-to-one basis. Participants were handed a written story. They were given 3 minutes to read the story the first time. When they finished the first reading, they were given 3 more minutes to read the story for a second and last time. They were not allowed to take any notes. Once they finished, the story was taken away. They then were asked to retell the story orally within 3 minutes. Their narratives were audio recorded and subsequently transcribed.

The second test administrated in the lab was the EI test and the procedure of administration was as described in the main study (see section 3.7). Similarly, the administration of the MLK test was as described in the main study.

3.6.1.1 Scoring, coding and data analysis

In order to analyze the data, specific scoring schemes were used for the tests. Then, scale reliability was checked, and descriptive statistics and correlational tests were conducted.

The scoring and coding for the measures BDS test, EI test, and ON test were as described in the main study (see section 3.8). For the latter, only the grammar points simple present tense, indefinite article, verb complements and simple past tense were included in the analysis because they reached the minimum of 3 occurrences as the threshold for giving a score for each grammar point. With respect to the LLAMA test, no scoring scheme was used because the computer did the scoring automatically after the completion of each subtest. The final score of each subtest was recorded for each participant.

The maximum possible score for the MLK test was 83. The MLK test was rated employing Roehr and Gánem-Gutiérrez's (2009b) scoring scheme (see Appendix I). For the correction section of the test, 1 point was awarded if the participant corrected the error in the sentence, or 0 points where the participant failed to correct the error. For the description/explanation section, these two were scored separately: 1 point for every adequate description (what form?), and 1 point for every adequate explanation (why this form?); 0 points were awarded if the participant was unable to describe/explain the correction of the error, or if the explanation/correction of the error was imprecise or incomplete, or if the participant translated or provided two options for the correction of the error. As for the scoring scheme for the second part of the test (i.e. the rule illustration section based on pedagogical grammar rules), 2 points were awarded for each item: 1 point for every correct sentence, and 1 point if the sentence included a clear context. Two points were awarded if the example sentence was correct and included a clear context, or 0 points were awarded if the participants failed to provide an example, or the example they provided was incorrect or was in the wrong context. If the participants used the target rule in a correct example, but without a clear context, 1 point was awarded. This scoring was modified in the main study (see section 3.8.1.2).

Prior to proceeding with the statistical tests, an internal reliability analysis was conducted for the MLK test for the cohort of participants as a whole and its subsections and for the EI test. The MLK test and the EI test showed overall good internal consistency (EI test, Cronbach's alpha = .88; MLK test, Cronbach's alpha = .71).

3.6.2 Learners' performance on the measures

Learners' performance on the LLAMA test is summarized in Table 3.15.

Table 3.15 Descriptive statistics for the LLAMA test

	N	Mean%	Max possible	Mean	SD	Min	Max
Level 5							
LLAMA in total	6	50	375	185.83	51.62	105	245
LLAMA_B (Vocabulary learning)	6	36	100	35.83	20.10	20	70
LLAMA_D (Sound recognition)	6	36	75	26.67	2.58	25	30
LLAMA_E (Sound-symbol correspondence)	6	83	100	83.33	24.22	40	100

LLAMA_F (Grammatical inferencing)	6	40	100	40.00	18.97	20	60
BDS in total	6	50	28	14.00	7.35	5	23
Level 7							
LLAMA in total	6	46	375	173.33	42.86	105	220
LLAMA_B (Vocabulary learning)	6	43	100	42.50	18.64	15	70
LLAMA_D (Sound recognition)	6	39	75	29.17	14.97	10	45
LLAMA_E (Sound-symbol correspondence)	6	75	100	75.00	25.88	30	100
LLAMA_F (Grammatical inferencing)	6	27	100	26.67	27.33	0	80
BDS in total	6	48	28	13.50	4.18	6	18
Level 9							
LLAMA in total	6	57	375	214.17	67.93	115	320
LLAMA_B (Vocabulary learning)	6	59	100	59.17	24.17	35	100
LLAMA_D (Sound recognition)	6	47	75	35.00	14.83	20	60
LLAMA_E (Sound-symbol correspondence)	6	83	100	83.33	19.66	50	100
LLAMA_F (Grammatical inferencing)	6	37	100	36.67	25.82	0	70
BDS in total	6	46	28	13.00	4.69	9	21

It is clear from the means in Table 3.15 that the group of learners in Level 9 outperformed the learners in the other two levels on every single subtest with the exception of LLAMA_E (sound-symbol correspondence) and LLAMA_F (grammatical inferencing); the group of Level 5 had the same mean for LLAMA_E and a higher mean for LLAMA_F than the other groups. Level 5 outperformed the group of Level 7 on the LLAMA_E test (sound-symbol correspondence) and the LLAMA_F test (grammatical inferencing). That participants scored lower or higher in each subtest indicates that the test is suitable for the purpose of the main study. The same applies for the results of the BDS test.

With regard to the measures of explicit and implicit knowledge of selected grammatical structures, descriptive statistics on participants' scores were conducted for both tests of implicit knowledge (i.e. the oral narrative test and the EI test) and for the test of explicit knowledge (i.e. the MLK test). Table 3.16 shows the results for both measures of implicit knowledge.

Table 3.16 Descriptive statistics for learners' performance on the EI and ON test

EI test					ON test				
Variable	Overall	Level 5	Level 7	Level 9	Variable	Overall	Level 5	Level 7	Level 9
Mean %	38	35	37	42	Mean %	63	55	69	65
Mean	35	33	34	39	Mean	63	55	69	65
SD	10.15	5.10	10.28	14.05	SD	18.90	23.78	14.88	17.52
Min	22	25	25	22	Min	n/a	n/a	n/a	n/a
Max	56	39	53	56	Max	n/a	n/a	n/a	n/a

Table 3.16 indicates that the learners obtained an overall higher score on the ON test than on the EI test. A similar difference in performance is reflected by group level. The difference in results between the two measures can potentially be attributed to the number of grammar points measured in each test.

The descriptive statistics in Table 3.17 show students' performance on the MLK test for each of the groups tested.

Table 3.17 Descriptive statistics of each level group for the MLK test

MLK test overall score	N	Mean%	Max possible	Mean	SD	Min	Max
MLK in total	6	66	83	54.00	7.78	36	68
MLK correction	6	86	17	14.56	1.62	11	16
MLK description/explanation	6	42	34	14.39	4.54	5	24
MLK rule illustration	6	80	32	25.44	3.11	19	30

Level 5							
MLK in total	6	61	83	51.00	10.66	36	68
MLK correction	6	80	17	13.67	2.16	11	16
MLK description/explanation	6	40	34	13.17	6.43	5	24
MLK rule illustration	6	76	32	24.17	3.12	20	29
Level 7							
MLK in total	6	66	83	55.17	7.46	49	68
MLK correction	6	90	17	15.33	1.21	13	16
MLK description/explanation	6	43	34	14.50	4.13	10	22
MLK rule illustration	6	79	32	25.33	4.08	19	30
Level 9							
MLK in total	6	69	83	57.00	3.74	53	61
MLK correction	6	86	17	14.67	1.03	13	16
MLK description/explanation	6	46	34	15.50	2.88	12	20
MLK rule illustration	6	84	32	26.83	1.47	25	29

As can be seen in Table 3.17, the higher the level of English of each group of participants the better the performance was with the exception of the students' performance of Level 7 for correction of sentences, in which they scored the highest of the three groups.

Correlations were calculated for the mean scores of every measure for the sample as a whole. Table 3.18 displays the correlations between the participants' performance on the measures of implicit and explicit knowledge, language aptitude and working memory.

Table 3.18 Correlations (Pearson *r*): Tests of MLK, EI, ON, LLAMA and BDS

	ON test	MLK test	LLAMA test	BDS test
EI test	.09 *p* = .71	.48* *p* = .04	.04 *p* = .86	.08 *p* = .76
ON test		.134 *p* = .60	-.16 *p* = .53	.16 *p* = .53
MLK test			.18 *p* = .47	-.02 *p* = .95
LLAMA test				.34 *p* = .17

* Correlation is significant at the 0.05 level (2-tailed)

No significant correlations were found between learners' language aptitude and implicit and explicit knowledge nor between learners' working memory and implicit and explicit knowledge, with the exception of a significant correlation between implicit and explicit knowledge (r = .45, p < .05). A further correlational analysis was run including the MLK subtests and the LLAMA subtests; moderate to strong correlations were found between the MLK subtests and similar correlations were obtained between the LLAMA subtests. No further correlations were calculated by level group due to the small number of participants.

To summarize, the findings obtained in this second pilot study suggested that the difficulty judgement questionnaires should be administered to all participants (teachers and learners) of the main study because a different categorization of grammar points might result, and in consequence, different results may be obtained for each implicit and explicit measure. The results of the EI test and the ON test suggested that these two implicit scores should be combined in the main study. Also, as suggested in the first pilot study, the grammar point prepositions + verbs must be substituted by the grammar point plural of nouns. The results in this second pilot study also suggest that the implicit and explicit measures are suitable to be used in a larger sample of participants. The main study addressed these conditions by recruiting a larger sample of learner participants (N = 90) and teacher participants (N = 26), and by combining the scores of the implicit measures.

3.7 Main study procedure

All participants agreed to participate in this study, and they received and signed a consent form before carrying out the completion of questionnaires and tests (see Appendix H). The order of administration for the learners was as follows: background information questionnaire, backward digit span test (BDS test), language aptitude test (LLAMA test), elicited imitation test (EI test), oral narrative test (ON test), metalinguistic knowledge test (MLK test), and difficulty judgement questionnaire, as summarized in Table 3.19.

Table 3.19 Summary of procedures of instruments for learners

	Order of administration of instruments and duration of data collection						
	Background questionnaire	**Backward digit span test**	**LLAMA test**	**Oral narrative test**	**Elicited imitation test**	**Metalinguistic knowledge test**	**Difficulty judgement questionnaire**
Location	Classroom	Quiet room	Lab	Quiet room	Lab	Classroom	Classroom
Timing	15 minutes	10 minutes	60 minutes	10 minutes	70 minutes	70 minutes	20 minutes
Number of sessions	1 session	1 session	2 sessions	1 session	1 or 2 sessions	2 sessions	1 session
Data collection	1 week/2 Saturdays	1 week/2 Saturdays	1 week/1 Saturday	1 week/2 Saturdays	1 week/1 Saturday	1 week/2 Saturdays	1 week/2 Saturdays

The teachers completed the difficulty judgement questionnaires in their own time, and all teachers returned their completed questionnaires.

As can be seen in Table 3.19, the first measure administered to learners was the background information questionnaire which they completed in one session before a regular class ended. Following the administration of these questionnaires, learners completed the BDS test on a one-to-one basis with the researcher in a quiet room. This was followed by the LLAMA test which took place in a lab supervised by the researcher in a single session. The EI test also took place in a lab, and it was completed in two sessions during weekday classes and in one session on Saturday classes; during weekday classes in the first session learners completed the training section of the test and in the second session they completed the test proper. The oral narrative test took place on a one-to-one basis with the researcher in a quiet room. The last two measures administered were the MLK test and the difficulty judgement questionnaire. Student participants completed the former in two sessions and the latter in one session in both weekdays and Saturday classes (completion of MLK test Part I in one session and part II in the following session; 35 minutes approximately for each session).

For the backward digit span test, the participants first completed a short training section with two sets of three and four sequences of digits, and then they did the test. The researcher read each sequence of digits to each participant in a monotone voice, one second apart. Student participants had to repeat the numbers backwards. The same sequence of numbers was read again, and learners had to repeat the numbers one more time in reverse order. All instructions for this test were given in Spanish (i.e., the learners' L1), and the test was completed in Spanish as well. All learners' answers were scored during the test.

For the ON test, participants were given a written story, and they were told that they were to read it twice. They were told that they would need to read it carefully because they were going to be asked to retell the story in as much detail as possible. They were not allowed to take any notes during this test. Three minutes were allotted to each reading and once they finished reading the story for the second time, the story was removed. They were then given three minutes to tell the story and they were told that they needed to keep as close to the original as possible in retelling it, using direct speech where appropriate. Their stories were audio recorded and subsequently transcribed.

With regard to the LLAMA test, each subtest is composed of two phases: (1) a timed study phase, with the exception of LLAMA_D, and (2) a testing phase, which is not timed (instructions were given in Spanish). Learners were asked to follow the following instructions for each subtest:

LLAMA_B: A vocabulary learning task

(1) Phase one: The main panel displays twenty objects. The task is to learn as many words as possible by associating each of them with a target object in the time available. For demonstration purposes, participants had 20 seconds to explore phase one of the program. When the program started, they proceeded to click on one of the buttons in the main panel, and the program displayed, in the centre of the panel, the name of the object that was clicked. Participants then proceeded to click the objects as many times as they wanted, but they could not take notes as they worked. The clock in the centre of the main panel showed them how much time was left to complete the task. When time was up, the program warned them by playing a bleep sound, and all the main buttons were deactivated.

(2) Phase two (testing phase): When the program started, participants had 2 minutes to learn the words in phase 1. When they finished, they proceeded to the testing phase. Once the test sequence was started, the program displayed the name of an object in its central panel, and waited for them to identify the correct object by clicking on it. The program gave them feedback in the form of a *ding* for a correct answer, and a *bleep* for an incorrect answer. Participants scored one point for each object that was correctly identified by its name. They were not penalized for guessing. The screen displayed the score as participants worked through the test.

LLAMA_D: A sound recognition task

(1) Phase one: This subtest does not present a timed phase. There is no time for demonstration purposes.

(2) Phase two (testing phase): The task is to listen carefully to 10 words in a language that is unfamiliar to participants. When the program started, participants had to listen carefully to the 10 words. When the program finished playing all the new sounds, participants heard a *bleep* sound, and the buttons on the main panel were activated. Participants could then proceed with the testing phase. They heard one word at a time, and if they thought it was a word that they had already heard, they clicked the ☺ button. If they thought it was a word that they had not heard before, they clicked the 😐 button. The program gave feedback in the form of a *ding* for a correct answer, and a *bleep* for an incorrect answer. Participants scored points every time they were right, but they lost points if they made a wrong judgement.

LLAMA_E: A sound-symbol correspondence task

(1) Phase one: The main panel displays twenty-four spellings (see Figure 2.2). The task is to use the time available to learn how the spelling system of this language works. For demonstration purposes, participants had 20 seconds to explore phase one of the program. When the program started, they clicked on one of the buttons in the main panel, and the program played a short sound file. The text on the button told participants how that particular sound was written in the language. The clock in the centre of the main panel showed how much time they had left to complete the task. When the time was up, the program warned them by playing a bleep sound, and all the main buttons were deactivated.

(2) Phase two (testing phase): When the program started, participants had 2 minutes to listen to the sound sequences as many times as they liked, and they could take any written notes that they needed. Once the test sequence started, the program played a new word. At the same time, it displayed two possible spellings for this word. One spelling was correct, and the other one was wrong. Participants proceeded to click on the spelling that they thought was correct. The program gave them feedback in the form of a *ding* for a correct answer, and a *bleep* for an

incorrect answer. Participants scored points for every item they got right, and lost points for incorrect answers. The screen displayed the score as they worked through the test.

LLAMA_F: A grammatical inferencing task

(1) Phase one: The main panel displays twenty small buttons. The task is to learn as much as possible about a new language in the time available. For demonstration purposes, participants had 60 seconds to explore the program. When the program started, they clicked on one of the small buttons in the main panel, and the program displayed a picture and a sentence that described it in the square on the right of the panel. They then proceeded to click the small buttons as many times as they liked, and they could take any written notes that they needed. The clock on the right-hand side margin of the main panel showed how much time they had left to complete the task. When the time was up, the program warned them by playing a bleep sound, and all the main buttons were deactivated.

(2) Phase two (testing phase): When the program started, participants had 5 minutes to analyze as many sentences as they could in accordance with each picture displayed, and they could take any written notes that they needed. Once the test sequence started, the program displayed a picture and two sentences. One sentence was grammatically correct, while the other contained a major grammar error. Participants then proceeded to click on the sentence that they thought was correct. The program gave them feedback in the form of a *ding* for a correct answer, and a *bleep* for a wrong answer. Participants scored points for every item they got right, and lost points for incorrect answers. The screen displayed the score as they worked through the test.

The second test administered in the lab was the EI test. Participants first did the training section of the test which consists of six example sentences. The informants heard each statement twice which the researcher recorded on a digital audio recorder. They then were allowed three seconds to judge whether the statements were “true”, “not true”, or “not sure” and circle their option on an answer sheet. After this they were given 8, 10, and 12 seconds, depending on the length of each sentence, to repeat the statement they had heard and record their utterances in the Audacity program version 2.0.5. They were given these seconds because in the software they had to click the recording option, click on the stop option, and click at the end of the track of that recording to be ready to begin with the following recording; the researcher regulated the timing of the presentation of the statements. The time of administration for this test took 70 minutes approximately.

The MLK test (both parts) was administered in the classroom in two sessions (30 to 40 minutes each session approximately) where participants had their English classes. They completed Part I of the MLK test in one session, and Part II in another session. They were explicitly told to write complete responses for the description/explanation and rule illustration section. For the description/explanation section they were at liberty to use either Spanish or English in their responses.

The difficulty judgement questionnaire was completed at the end of a class. Students were instructed to judge the learning difficulty of each grammar point according to their learning experience by ticking a 5-point scale ranging from very easy to very difficult.

3.8 Scoring, coding and data analysis

In order to analyze the data, specific scoring schemes were used for the tests. Then scale reliability and normality of distribution were checked, and descriptive statistics, correlational tests, analysis of variance tests (ANOVA) and regression analyses were conducted. ANOVA tests were run on a number of variables of the MLK and EI test despite a non-normal distribution of data in some cases. The analyses were conducted taking into account that ANOVA is robust to the violation of the assumption of normality when the groups being compared have equal sizes (see Field, 2013, pp. 444-5 for detailed discussion).

3.8.1 Scoring and coding

The scoring and coding of the LLAMA test and the BDS test are discussed first followed by the MLK test, EI test, ON test and the assessment of learners' speaking performance on the oral narrative test.

3.8.1.1 Language aptitude test and backward digit span test

With regard to the LLAMA test, the computer did the scoring automatically after the completion of each subtest. The scores for all subtests range between 0 to 100 except for the LLAMA D subtest, which the score ranges between 0 to 75. The final score of each subtest was recorded for each participant.

Participants' backward digit span was calculated as the highest number of digits the participants were able to repeat correctly at least two times out of four sequences within a set (Kormos & Sáfár, 2008). That is, if learners repeated the first sequence of three digits correctly in the first or second trial, they scored 1 point, but if they wrongly repeated the sequence in both trials, 0 points were scored. They continued with the next sequence and the procedure for scoring was the same. In the case where the participants successfully repeated two sequences from a set in the first trial, they scored 4 points and the other two sequences were skipped. When the participants repeated two consecutive sequences incorrectly, the test was terminated. The maximum possible score was 28 points. An example is provided below:

Level 5: S1

Example:

<table>
<tr><th>Sets</th><th>Sequences</th><th>Trial 1</th><th>Trial 2</th></tr>
<tr><td rowspan="4">1st set
(3 digits)</td><td>5-8-2</td><td>*1</td><td></td></tr>
<tr><td>6-9-4</td><td>1</td><td></td></tr>
<tr><td>1-4-8</td><td></td><td></td></tr>
<tr><td>2-7-6</td><td></td><td></td></tr>
<tr><td rowspan="3">2nd set
(4 digits)</td><td>6-4-3-9</td><td>1</td><td></td></tr>
<tr><td>7-2-8-6</td><td>0</td><td>1</td></tr>
<tr><td>9-6-2-5</td><td>0</td><td>1</td></tr>
</table>

	7-4-9-1	0	0
3rd set (5 digits)	4-2-7-3-1	0	**0
	7-5-8-3-6		
	6-4-7-8-1		
	9-6-2-7-4		

*If sequence repeated correctly in reverse order.
**Test was terminated because of failure on two consecutive sequences.

3.8.1.2 Difficulty judgement questionnaires and metalinguistic knowledge test

Concerning the teachers' and learners' difficulty judgement questionnaires, teachers' and learners' responses were converted to the following scores: very easy = 1, easy = 2, moderate = 3, difficult = 4 and very difficult = 5.

The MLK test was rated employing Roehr and Gánem-Gutiérrez's (2009b) scoring scheme (see Appendix I). The test was scored dichotomously, and the maximum possible score was 52 points. In the first part, 1 point was awarded if the participant corrected the error in the sentence, or 0 points where the participant failed to correct the error for most grammar points represented by one test item. As for the five grammar points represented by two test items (i.e. comparative adjectives, infinitives and gerunds, many vs. much, since/for, and yes/no questions) involving a two-part pedagogical rule, 0.5 point was awarded if the participant corrected the error in the sentence (see Appendix J). For example, the following responses were scored correct.

Part I
Item 12: Participant 7
Target structure: Yes/no questions
Stimulus: **Do** Pedro work late?
Response: does
Score: 0.5 point

Item 8: Participant 1
Target structure: Indefinite article
Stimulus: She bought **the** new house.
Response: a

Score: 1 point

Additionally, 1 point was awarded for every adequate description (what form?), and 1 point for every adequate explanation (why this form?), except for the five grammar points represented by two test items for which 0.5 point was awarded. For example, the following response was scored correct (1 or 0.5 point for the description and 1 or 0.5 point for the explanation).

Item 3: Participant 17
Target structure: 3rd person –*s*
Stimulus: Sara **cook** every day.
Response: (description/explanation): When we have a sentence in simple present and we are talking about a third person we have to add an "s" or "es" in the verb.
Score: 2 points

Description (1 point): When we have a sentence in simple present
Explanation (1 point): we are talking about a third person we have to add an "s" or "es" in the verb.

Item 6: Participant 17
Target structure: Comparative adjectives
Stimulus: Mike is **more tall** than Joe.
Response: (description/explanation): *Los adjetivos de una sílaba incluyen el comparativo en la terminación "-er".* (Comparative adjectives of one syllable include the –*er* ending).
Score: 1 point

Description (0.5 point): Comparative adjectives of one syllable
Explanation: (0.5 point): include the –*er* ending

Zero points were awarded if the participant was unable to describe or explain the correction of the error. For example, the following response was scored as incorrect (0 points for the explanation and 0 points for the description).

Item 1: Participant 13
Target structure: Simple past tense (-*ed* form)

Stimulus: When he finished his homework, he **watch** a movie.

Response: (description/explanation): *Is past progressive because se interrumpe la acción con el pasado simple* (It is past progressive because the action is interrupted with the simple past).

Description (0 points): It is past progressive

Explanation (0 points): because the action is interrupted with the simple past

Zero points were awarded if the explanation or description of the error was imprecise or incomplete. For example, in the following response 1 point was awarded for the description and 0 points for the explanation.

Item 8: Participant 4

Target structure: Indefinite article

Stimulus: She bought **the** new house.

Response: (Description/explanation): *Debe utilizarse artículo definido* (a definite article should be used)

Description (1 point): an indefinite article should be used

Explanation (incomplete, 0 points):

Zero points were awarded if the participant translated the stimulus sentence instead of describing or explaining the correction of the error. For example, the following response was scored as incorrect.

Item 7: Participant 10

Target structure: Verb complements

Stimulus: They finished **to build** the house.

Response: (Description/explanation): *Ellos ya terminaron de construirla* (They finished building it).

Zero points were awarded if the participant provided two options for the explanation or description of the error. For example, the following response was scored as incorrect.

Item 10: Participant 14

Target structure: Many vs. much

Stimulus: I have **many** money.

Response: (Description/explanation): *Porque es contable o no contable* (Because it is countable or uncountable).

In part II, (i.e. the rule illustration section based on pedagogical grammar rules), 1 point was awarded if the example sentence was correct or 0 points if the example sentence was incorrect. For sentences illustrating six pedagogical grammatical rules (simple past tense, 2^{nd} conditional, comparative adjectives, infinitives and gerunds, many vs. much, and since/for), they were awarded 0.5 or 1 point as follows: In the case of the grammar point simple past, 0.5 point was awarded to the correct example sentence and 0.5 point to the context; for the grammar point 2^{nd} conditional, 0.5 point was awarded for each correct clause; for the rest of the grammar points that require two sentences, 0.5 point was awarded to each correct sentence, or 0 points were awarded if the participants failed to provide an example, or the example they provided was incorrect or was in the wrong context. What follows is an example for each of these situations.

Part II

Grammar points requiring two sentences

Item 4 and 5: Participant 6

Item 4: Targeted grammar point: Comparative adjectives

Rule: When a comparative is formed for a one-syllable adjective, *-er* is added.

Example sentence1: My cell phone is big than his head.

Score: 0 points

Item 5: Rule: When a comparative is formed for an adjective with two or more syllables, *more* is placed in front.

Example sentence2: My car is more fantastic than his truck.

Score: 0.5 point

Grammar point requiring two clauses

Item 2: Participant 13

Targeted grammar point: 2^{nd} conditional

Rule: When an unreal/hypothetical situation is being expressed, the 2^{nd} conditional comprising an *if*-clause with a past tense verb and a main clause with *would* + *verb* is used.

Example sentence: If I won the lottery I would buy a car.

Score: 1 point (0.5 point for each clause)

With regard to the grammar point simple past tense, if the participants gave a completely correct example with an appropriate context (where necessary), 0.5 point was awarded to the correct example and 0.5 point to the appropriate context. However, if the participants used the target rule in a correct example, but without a clear context (when necessary) or no context provided 0.5 point was awarded to the correct example and 0 points to the unclear context or no context provided. If the participants provided an incorrect example including a clear context, 0.5 point was awarded to the context and 0 points to the incorrect example. In the case where the participants gave an incorrect example without a clear context (where necessary), or no context provided, 0 points were awarded.

Example sentence with an appropriate context

Item 1: Participant 2

Example sentence: He played in the park yesterday.

Score: example sentence 0.5 point, context 0.5 point.

Example sentence without a clear context

Item 1: Participant 13

Target structure: Simple past tense (-*ed* form)

Rule: When a finished action or event in the past is being expressed, the simple past tense is required. (Please use a regular verb).

Example sentence: Jorge played basketball.

Score: example sentence 0.5 point, context (not provided) 0 points.

Incorrect example sentence with clear context
Item 1: Participant 7
Target structure: Simple past tense (-*ed* form)
Example sentence: I study for the exam yesterday.
Score: example sentence 0 points, context 0.5 point.

For the remaining grammar points, 1 point was awarded if the example sentence was correct or 0 points if the example sentence was incorrect. All spelling mistakes in the example sentences were ignored. To ensure inter-rater reliability of the scoring on the MLK test, a second rater was asked to mark a randomly selected sub-sample of scripts (20%) after I had scored the items of this test. The Pearson's *r* correlation coefficient of inter-rater reliability for the two raters was high for the MLK test for each group of learners: Level 5 (r = .977, p < .001), Level 7 (r = .995, p < .001), and Level 9 (r = .928, p < .01). Cases of disagreement were discussed until a resolution was reached.

3.8.1.3 Elicited imitation test and oral narrative test

On the EI test, student participants' responses were scored according to Erlam's (2006) three criteria: (1) obligatory occasion created – supplied; (2) obligatory occasion created – not supplied; and (3) no obligatory occasion created.

For the first criterion, a correct response was awarded 1 point irrespective of lexical accuracy or any grammatical errors on structures that did not pertain to the targeted grammar point; for example, the following responses (the asterisk "*" indicates an incorrect stimulus) were scored correct.

Item 2: Participant 5
Target structure: Verb complements (verb + infinitive)
Stimulus: Mexican people **want to keep** their country clean and green.
Response: Mexican people **want to see** their country clean and green.

Item 14: Participant 2
Target structure: Comparative adjectives
Stimulus: *It is **more harder** to learn Japanese than to learn English.
Response: It is more difficult to learn Japanese than to learn English.

For the second criterion, a response was awarded 0 points where the participant created an obligatory occasion for use of the target structure but used it incorrectly. For example, the following responses were scored as incorrect.

Item 7: Participant 9
Target structure: Dative alternation
Stimulus: People should report a car accident **to** the police.
Response: People should report a car accident at the police.

Item 54: Participant 54
Target structure: Verb complements
Stimulus: *Most young people enjoy **to listen** to music.
Response: Most young people enjoy listen to music.

For the final criterion, 0 points were awarded to participants where they did not create an obligatory occasion for use of the grammar point. This includes responses where the participants did not attempt the section of the sentence that contained the grammar point. The maximum possible score was 93 points. For example, the following response was scored as incorrect.

Item 61: Participant 10
Target structure: Relative clauses
Stimulus: The two sports that most people watch are soccer and baseball.
Response: People watch soccer and baseball.

Item 8: Participant 6
Target structure: Since/For
Stimulus: *People have used computers **since** many years.

Response: People have used computers.

Regarding the oral narrative test, an obligatory occasion analysis (Brown, 1973; see Ellis & Barkhuizen, 2005; Gass & Selinker, 2008) was conducted to account for the suppliance in obligatory contexts (SOC) of the targeted grammar points. The rationale for using the SOC analysis instead of a target-like use analysis (Pica, 1984) is that student participants rarely supplied grammar points in non-obligatory contexts. The programs CHAT and CLAN of the Child Language Data Exchange System (CHILDES) (http://childes.talkbank.org) were used for the transcription, coding (CHAT) and analysis (CLAN) of each participant's recording on their retelling of the short story. Grammatical accuracy was measured on the basis of whether learners were accurate or not in producing the target grammar points in obligatory contexts. Thus, the accuracy score on any target grammar point would be calculated as follows: (accuracy = [N of correct occurrences x 100] / N of all occurrences). A minimum of 3 occurrences was used as the threshold for giving a score for each grammar point. The maximum possible score for each grammar point is thus 100%.

Student participants' responses were scored according to four criteria: If a grammar point was correctly supplied, 1 point was awarded.

Level 5: Participant 5
Target structure: 3rd person *–s*
Stimulus: He **walks** to the store and **buys** a newspaper.
Response: He walks to the store and buys the newspaper.

For the second criterion, if a different word or phrase was used from that of the original story but the message was correctly expressed, 1 point was awarded. For example, the following response was scored correct.

Level 9: Participant 19
Target structure: Simple past tense (-*ed* form)
Stimulus: Mr. Garcia **accepted** the five million dollars.
Response: Mr. Garcia received the money.

For the third criterion, if a word or a phrase was repeated immediately after it had been produced, it was counted only once. For example, the following response was awarded 1 point.

Level 7: Participant 2
Target structure: Verb complements
Stimulus: I want **to give** some money to you.
Response: He tried to help [/] to help him.

For the final criterion, an incorrect response was awarded 0 points where the participant created an obligatory occasion for use of the target grammar point but used it incorrectly. For example, the following response was scored as incorrect.

Level 5: Participant 1
Target structure: 3rd person –*s*
Stimulus: He **gets** up at 6:30 am.
Response: He get up at 6:30 am.

With regard to the proficiency level of English of each learner, I used the IELTS Speaking Band Descriptors. These band descriptors describe performance in four categories:

- fluency and coherence
- lexical resource
- grammatical range and accuracy
- pronunciation

Given that on the oral narrative task learners have to retell the short story, there is no opportunity for them to develop the topic, or to use a broad lexical range according to what is demanded by the band descriptors in the category of lexical resource, therefore, this criterion was precluded.

Likewise, the band descriptors about developing a topic and about the use of a range of connectives and discourse markers were not included as a holistic evaluative part of each band descriptor that list them under the criterion of fluency and coherence. The criteria of grammatical range and accuracy and pronunciation were not modified. Therefore, three criteria were used to assess the learners' speaking performance on the oral narrative test:

- fluency and coherence
- grammatical range and accuracy
- pronunciation

I first assessed the learners' speaking performance on the oral narrative test according to these criteria, and then I asked 5 out of the 26 teachers who completed the difficulty judgement questionnaire to participate in assessing the learners' speaking performance on the same test and using the same assessment criteria. The criteria were tried out as follows:

1) Six recordings were selected for training purposes; the first two recordings from each course level (i.e. two from Level 5, two from Level 7, and two from Level 9).
2) The six recordings were rated following the IELTS Speaking Band Descriptors on the basis of each of the three criteria. The results were reported as band scores on a scale from 1 (the lowest) to 9 (the highest) (see Table 3.20 - https://www.ielts.org/about-the-test/how-ielts-is-scored).

Table 3.20 The IELTS scale

Band score	Skill level	Description
9	Expert user	The test taker has fully operational command of the language. Their use of English is appropriate, accurate and fluent, and shows complete understanding.
8	Very good user	The test taker has fully operational command of the language with only occasional unsystematic inaccuracies and inappropriate usage. They may misunderstand some things.
7	Good user	The test taker has operational command of the language, though with occasional inaccuracies, inappropriate usage and misunderstandings in some situations. They generally handle complex language well and

		understand detailed reasoning.
6	Competent user	The test taker has an effective command of the language despite some inaccuracies, inappropriate usage and misunderstandings. They can use and understand fairly complex language, particularly in familiar situations.
5	Modest user	The test taker has a partial command of the language and copes with overall meaning in most situations, although they are likely to make many mistakes. They should be able to handle basic communication in their own field.
4	Limited user	The test taker's basic competence is limited to familiar situations. They frequently show problems in understanding and expression. They are not able to use complex language.
3	Extremely limited user	The test taker conveys and understands only general meaning in very familiar situations. There are frequent breakdowns in communication.
2	Intermittent user	The test taker has great difficulty understanding spoken and written English.
1	Non-user	The test taker has no ability to use the language except a few isolated words.
0	Did not attempt the test	The test taker did not answer the questions.

3) A deviation of up to 2 points between teachers was regarded as acceptable for each criterion.
4) Then, teachers (including myself) were randomly assigned 60 oral narratives; consequently each recording was rated by 4 different teachers.
5) The mean score was then used to determine each learner's proficiency level.

3.9 Data analysis

Before proceeding with the statistical tests, an internal reliability analysis was conducted for the MLK test as a whole and its subsections, the EI test, and the BDS test[4]. The normality of distribution for the whole sample of participants (N = 90) on the EI, ON, MLK, LLAMA, BDS tests and L2 proficiency was checked, as well as the normality of distribution for the implicit knowledge (i.e. EI/ON combined scores) and explicit knowledge (i.e. MLK scores) scores. Furthermore, a comparative analysis was conducted between learner and teacher judgements on the degree of learning difficulty for each of the 13 targeted grammar points.

[4] The LLAMA test does not provide itemized data (Granena, 2013)

3.9.1 Reliability and normality analysis

Table 3.21 presents the overall Cronbach's alpha for the tests of MLK, EI and BDS.

Table 3.21 Cronbach's alpha for the tests of MLK, EI and BDS

	N	Number of items	Cronbach's alpha
EI total	90	93	.92
BDS in total	90	28	.90
MLK total	90	52	.86

As can be seen in Table 3.21, all tests show good reliability (Cohen et al., 2011) if taken as a whole.

To ensure inter-rater reliability of the scoring on the MLK test, a second rater was asked to mark a randomly selected sub-sample of scripts (20%) after I had scored the test. The Pearson's *r* correlation coefficient of inter-rater reliability for the two raters was high (r = .885, p < .001). Cases of disagreement were discussed until a resolution was reached.

The analysis for normality presented the following results for all instruments as shown in Table 3.22.

Table 3.22 K-S tests of normality for the whole sample of participants (N = 90) on the EI, ON, MLK, LLAMA, BDS tests and L2 proficiency

Tests	Sig.
EI test	.641
ON test	.312
MLK test	.157
MLK correction	.001
MLK description/explanation	.191
MLK rule illustration	.509
LLAMA test	.918
LLAMA B (Vocabulary learning)	.514
LLAMA D (Sound recognition)	.377
LLAMA E (Sound-symbol correspondence)	.003

LLAMA F (Grammatical inferencing)	.084
BDS test	.026
L2 proficiency	.033

As shown in Table 3.22, the non-significant *p*-values ($p > .05$) indicate a normal distribution for the overall scores on the EI, ON, MLK, and LLAMA test except for the BDS test and L2 proficiency. It should be noted that Table 3.22 also includes the normality of distribution of scores for the LLAMA subtests and MLK test subsections.

The assumption of normal distribution was met for all MLK test subsections except for the MLK correction section, but taking into account that a Pearson product moment correlation is a robust measure that can cope with some violations (see Pallant, 2010; Norman, 2010; see also Trafimov & Marks, 2015), the correlation was conducted as shown in Table 3.23.

Table 3.23 Correlation between the MLK test and each of its subsections

	MLK correction	MLK description/ explanation	MLK rule/ illustration
MLK test	.739**	.963**	.664**

**. Correlation is significant at the 0.01 (2-tailed).

As shown in Table 3.23, the MLK test correlated strongly with each of its subsections (see Cohen et al., 2011, p. 636 for strength of correlation coefficients), which indicates that each of the subsections represents a coherent subordinate measure of metalinguistic knowledge. This relationship was highest between the MLK test and its subsection description/explanation where the learners had to describe or explain the error of each sentence, suggesting that this subsection represented the MLK test the most.

In order to check the distribution of the EI test scores, the implicit knowledge (i.e. EI/ON combined scores) scores (see Table 3.22) and explicit knowledge (i.e. MLK scores) scores for each group of participants (Level 5, Level 7, Level 9), one-sample Kolmogorov-Smirnov tests were conducted. The tests revealed that the distribution of scores did not differ significantly from a normal distribution on these scores. Table 3.24 illustrates these normal distributions.

Table 3.24 K-S tests of normality for each group of student participants (levels 5, 7, and 9)

EI test	Sig.
Level 5	.968
Level 7	.872
Level 9	.932
Implicit knowledge	
Level 5	.814
Level 7	.855
Level 9	.572
Explicit knowledge	
Level 5	.403
Level 7	.543
Level 9	.868

3.10 Overall summary

This chapter describes the research design, the learner participants (N = 90) and teacher participants (N = 26), the difficulty judgement questionnaires, the explicit measure (MLK test), the implicit measures (EI and ON test), the language aptitude test and the working memory capacity test, the piloting of the measures, and the scoring, coding and data analysis of the measures.

The group of teacher participants comprised 26 teachers teaching English at different levels on the English Extension Program in the same university the learners were studying English.

The learners recruited in the present study were learning English as a foreign language in a so-called English Extension Program of a university in Mexico following a focused-form instruction. The learner participants were in three different groups: one group at Level 5 (intermediate, *n* = 30), one group at Level 7 (upper-intermediate *n* = 30), and one group at Level 9 (advanced, *n* = 30).

The learners' explicit knowledge of 13 targeted grammar points, perceived by learners and teachers as easy grammar points (plural of nouns, simple past tense (*-ed* form), modal verbs, *many* vs. *much*, comparative adjectives, *since* vs. *for*) and difficult grammar points (indefinite article, simple present tense (3rd person –*s*), verb complements, second conditional (*if*-clauses), *yes/no* questions, dative alternation, relative clauses) (see section 4.1), was operationalised by the MLK test. Learners' implicit knowledge was operationalised as the combination of the scores of the EI and ON test. In order to facilitate test design and subsequent scoring, the pedagogical grammar rules corresponding to the 13 grammar points were converted into the format "When form X occurs/function X is being expressed, form Y needs to be used" (Roehr, 2008; Thepseenu & Roehr, 2013; Ziętek & Roehr, 2011). The ON test contains eight of the targeted structures, a subset of the 13 grammar points included in the EI test and the MLK test. All tests were piloted prior to their administration in the main study. The tests were planned to be administered from less-to-more awareness of the grammar rules of the grammar points, that is, the order of administration was the ON test, the EI test, the MLK test, and the difficulty judgement questionnaires.

With regard to test reliability and normality of distribution, the EI, BDS, and MLK test showed good reliability; the LLAMA test does not provide itemized data (Granena, 2013) to run internal reliability analysis. The tests EI, ON, MLK, and LLAMA showed a normal distribution for the overall scores with the exception of the BDS test and L2 proficiency. The assumption of normal distribution was also met for all MLK test subsections and LLAMA subtests except for the MLK correction section and the LLAMA_E (Sound-symbol correspondence) subtest. Similarly, the analysis for normality for the whole sample of participants for the easy and difficult grammar points on the EI, ON, and MLK test was normal (see Table 4.6). Furthermore, it was argued that parametric measures were used to run correlations between variables despite the non-normal distribution of some measures.

CHAPTER FOUR: RESULTS

The results are presented in the order of the research questions of this study which are repeated in the following sections for convenience.

4.1 RQ1. How did the teacher and learner participants judge the targeted grammar points in terms of learning difficulty?

In order to address the issue of the difficulty judgement of grammar points, learner and teacher rankings of each grammar point (1 for "very easy" and 5 for "very difficult") are presented employing the scoring scheme discussed in the scoring and coding section. To make the interpretation of these values as to whether a construction is easy or difficult, the mean difficulty score was calculated for each grammar point. Table 4.1 shows the descriptive statistics for the learners' judgements of the difficulty of the grammar points, ordered from the easiest (i.e. lowest mean scores) to the most difficult (i.e. highest mean scores).

Table 4.1 Learner mean judgements on the difficulty of grammar points

Grammar points	Mean	SD
Simple past tense (-*ed* form)	1.50	.64
3rd person –*s*	1.51	.60
Plural of nouns	1.74	.71
Comparative adjectives	1.88	.78
Yes/*no* questions	1.89	.76
Modal verbs	1.96	.75
Many vs. much	2.17	.67
Since/for	2.24	.78
Indefinite article	2.33	.82
Second conditional (*if*-clause)	2.47	.93
Dative alternation	2.67	.96
Verb complements	2.82	.97
Relative clauses	2.92	1.01

The mean scores in Table 4.1 indicate that none of the grammar points was perceived as difficult or very difficult. The rating scale (1 – 5) was modified to include two main difficulty categories: easy and difficult. Table 4.2 shows how the 13 grammar points were classified accordingly.

Table 4.2 Difficulty of grammar points according to learner judgements

Easy (1.5 – 2.2)	**Difficult (2.3 – 2.9)**
1. Simple past tense (-*ed* form)	1. Indefinite article
2. 3rd person –*s*	2. Second conditional
3. Plural of nouns	3. Dative alternation
4. Comparative adjectives	4. Verb complements
5. *Yes/no* questions	5. Relative clauses
6. Modal verbs	
7. Many vs. much	
8. *Since/for*	

As can be seen in Table 4.2, eight grammar points were perceived as easy by the learners and five were rated as difficult. In comparison to other studies (Absi, 2014; Huang, 2012; Rodríguez & Roehr-Brackin, 2016; Thepseenu & Roehr, 2013), the learners' judgements of this study showed a similar tendency towards the easy side of learning difficulty despite the differences of age, language experience, school education and L1 of the participants from the other studies.

Statistical analyses were also run to examine the teachers' judgments. Table 4.3 shows the descriptive statistics for the teachers' judgements of the difficulty of the grammar points, ordered from the easiest (i.e. lowest mean scores) to the most difficult (i.e. highest mean scores).

Table 4.3 Teacher mean judgements on the difficulty of grammar points

Grammar points	**Mean**	**SD**
Plural of nouns	2.23	.77
Simple past tense (-*ed* form)	2.58	.75
Modal verbs	2.73	.78
Many vs much	2.88	.91

Comparative adjectives	2.92	.85
Since/for	3.04	.87
Indefinite article	3.15	1.01
3rd person *–s*	3.19	1.20
2nd conditional (*if*-clause)	3.23	.71
Yes/no questions	3.27	1.04
Dative alternation	3.62	.85
Verb complements	3.77	.86
Relative clauses	3.81	.57

The mean scores in Table 4.3 indicate that none of the grammar points was perceived as very easy or very difficult. Table 4.4 shows how the teachers' ratings were classified according to the two difficulty categories: easy and difficult.

Table 4.4 Difficulty of grammar points according to teacher judgements

Easy (2.2 – 3)	**Difficult (3.1 – 3.8)**
1. Plural of nouns	1. Indefinite article
2. Simple past tense (-*ed* form)	2. 3rd persons *–s*
3. Modal verbs	3. 2nd conditional (*if*-clause)
4. Many vs. much	4. Yes/no questions
5. Comparative adjectives	5. Dative alternation
6. Since/for	6. Verb complements
	7. Relative clauses

Table 4.4 indicates that teachers classified six grammar points as easy and seven as difficult. A comparison between Table 4.2 and 4.4 shows that both learners and teachers perceived the same grammar points as easy or difficult, with the exception of two grammar points: 3rd person *–s* and yes/no questions are grouped as easy by the learners and as difficult by the teachers. These two grammar points were assigned to the difficult category by the researcher for three reasons: (1) a Spearman's rank order correlation was run between the learners' and the teachers' mean

difficulty scores to find out whether there was any correlation. There was a significant and strong positive correlation between the two sets of mean scores (rho = .742, $p < .01$), (2) one study used the expert judgement of experienced L2 teachers to categorize a number of grammar points as easy or difficult (Robinson, 1996) and, (3) in a study that was conducted prior to this study (Rodríguez & Roehr-Brackin, 2016) where 30 learners, 3 applied linguists (including the researcher of this study), and 11 teachers made learning difficulty judgements on a number of grammar points; the teacher group was the only group who showed a tendency towards successful prediction of learners' performance on both explicit and implicit L2 measures (i.e. a metalinguistic knowledge test and an elicited imitation test on which the tests used in the present study were based). Table 4.5 shows the final classification of the grammar points.

Table 4.5 Difficulty of grammar points according to learner and teacher judgements

Easy	**Difficult**
1. Plural of nouns	1. Indefinite article
2. Simple past tense (-*ed* form)	2. 3rd person –*s* in the simple present tense
3. Modal verbs	3. 2nd conditional (*if*-clause)
4. Many vs. much	4. *Yes/no* questions
5. Comparative adjectives	5. Dative alternation
6. *Since/for*	6. Verb complements
	7. Relative clauses

The analysis for normality presented the following results for the easy and difficult constructions on the EI, ON, and MLK test as shown in Table 4.6.

Table 4.6 K-S tests of normality for the whole sample of participants (*N* = 90) for the easy and difficult grammar points on the EI, ON, and MLK test.

	Sig.
EI easy grammar points	.777
EI difficult grammar points	.692

ON easy grammar points	.197
ON difficult grammar points	.279
EI/ON combined scores (easy grammar points)	.840
EI/ON combined scores (difficult grammar points)	.388
MLK easy grammar points	.127
MLK difficult grammar points	.156

Table 4.6 shows that the distribution of data for the implicit measures (i.e. EI and ON tests), the implicit combined scores of the EI and ON tests and explicit measure (i.e. MLK test) for easy and difficult grammar points was normal.

4.2 RQ2a. What is the level of learners' explicit knowledge of the targeted grammar points as measured by the metalinguistic knowledge test?

With regard to RQ2a, descriptive statistics were calculated for the students' scores on the MLK test for the cohort as a whole and for each group of participants (levels 5, 7, 9). The descriptive statistics in Table 4.7 show the students' performance on the MLK test and its subsections for the sample as a whole.

Table 4.7 Descriptive statistics for the performance on the MLK test and its subsections

	N	Mean %	Max Possible	Mean	SD	Min	Max
MLK in total	90	67	52	35.06	6.75	20	47
MLK correction	90	86	13	11.21	1.59	6.5	13
MLK description/explanation	90	54	26	14.03	4.76	3	23
MLK rule illustration	90	75	13	9.78	1.57	5.5	13

The descriptive statistics in Table 4.7 show that, on average, the MLK test was moderately difficult for the sample of participants (mean percent score = 67). It can also be noticed that the participants generally found the description/explanation section most difficult, probably due to the explicit articulation of the underlying regularity (i.e. the pedagogical grammar rule); in contrast to this result, learners found the correction section easiest. As previously stated in the literature review, learners' performance on the correction section may be explained by the additional use of implicit knowledge (R. Ellis, 2004) that may have occurred on the correction task, namely, the context is already provided for each sentence and the learner needs to only identify the correct form; the correction section does not require verbalization of the underlying regularity of the grammar point.

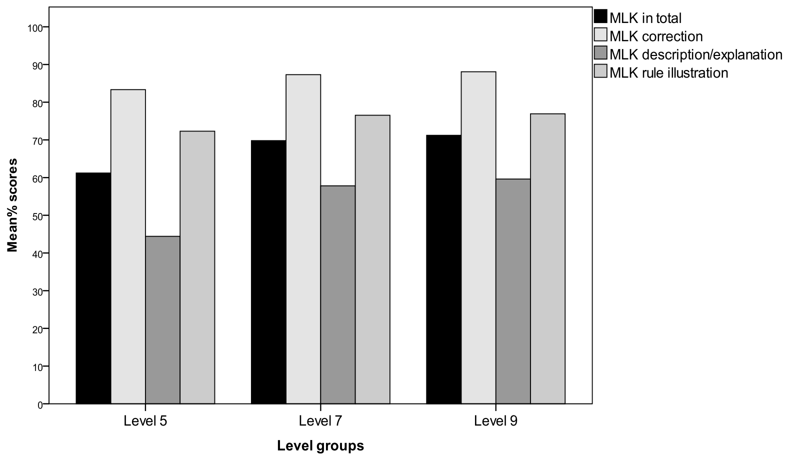
To proceed with the statistical analysis in relation to the grouping variable, descriptive statistics were calculated for each group of participants. Table 4.8 shows students' performance on the MLK test for each group tested.

Table 4.8 Descriptive statistics for the performance of each group of participants on the MLK test and its subsections

Level 5	**N**	**Mean%**	**Max possible**	**Mean**	**SD**	**Min**	**Max**
MLK in total	30	61	52	31.83	7.04	20	4
MLK correction	30	83	13	10.83	1.90	6.5	13
MLK description/explanation	30	44	26	11.55	4.71	3	19
MLK rule illustration	30	72	13	9.40	1.54	5.5	12
Level 7							
MLK in total	30	70	52	36.30	5.90	25	47
MLK correction	30	87	13	11.35	1.37	8	13
MLK description/explanation	30	58	26	15.03	4.48	5.5	23
MLK rule illustration	30	76	13	9.95	1.28	8	12
Level 9							
MLK in total	30	71	52	37.03	6.24	24	47
MLK correction	30	88	13	11.45	1.42	7.5	13
MLK description/explanation	30	60	26	15.50	4.20	4	21
MLK rule illustration	30	77	13	10	1.84	5.5	13

As can be seen in Table 4.8, the mean scores and variance of Level 7 and Level 9 are very similar. Level 5 scored lower in each subtest in comparison to the other two groups. In this group, all the scores are more spread out as indicated by their variances with the exception of the rule illustration section when compared to the variance of Level 9. It seems that the level of explicit knowledge of the last two groups is relatively "moderate" and for Level 5 is relatively "low" according to the overall score on the MLK test of each group. Figure 4.1 shows the students' performance on the MLK test for the three groups of participants.

Figure 4.1 Students' performance on the MLK test



Given that the assumptions of normality (see Table 3.22) and homogeneity of variances (Levene's test: $p > .05$) were met, a one-way ANOVA was conducted to find out whether the differences of the MLK total scores between groups were significant. The results indicate that there was a significant difference between groups ($F(2, 89) = 5.78$, $p = .004$). *Post hoc* tests (Tukey's test) show that there was a significant difference between Level 5 and 7 (mean difference = -4.467, 95% CI = -8.41, -.52, $p < .05$, representing a medium effect, $d = .63$[5]), Level 5 and 9 (mean difference = -5.200, 95% CI = -9.15, -1.25, $p < .05$, representing a medium effect, $d = .73$) but not between Level 7 and 9 (mean difference = -.733, 95% CI = -4.68, 3.21, $p =$

[5] Effect sizes; 0.2-0.5 = small, 0.5-0.8 = medium, > 0.8 = large (Cohen, 1988; Norris and Ortega, 2001)

.898). This indicates that there was a significant development in explicit knowledge between Level 5 and 7, but not between Level 7 and 9.

4.3 RQ2b. What is the level of learners' explicit knowledge of the easy and difficult grammar points as measured by the metalinguistic knowledge test?

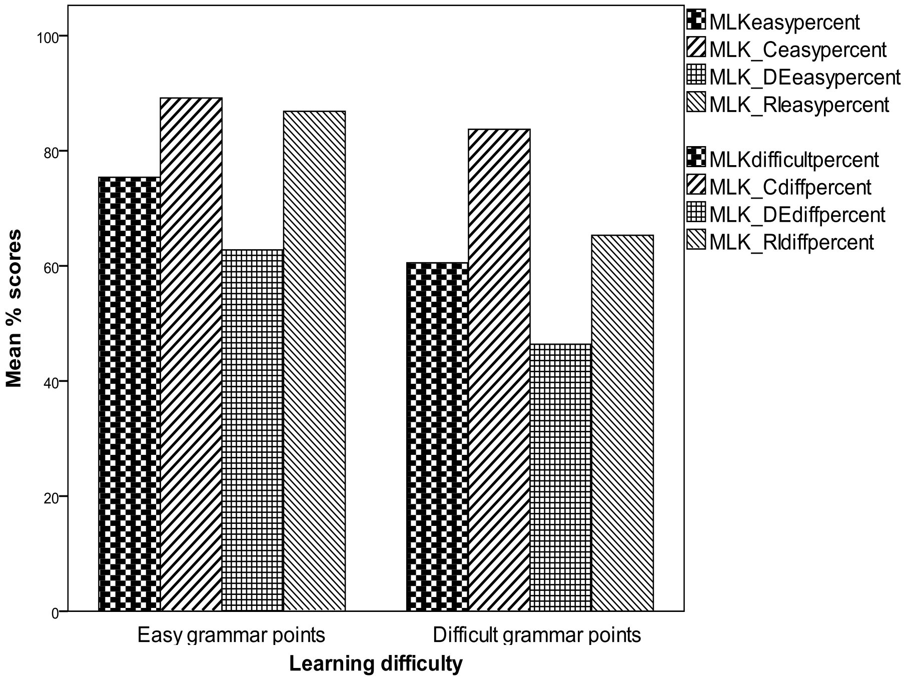
Descriptive statistics were calculated in accordance with learners' performance on the MLK test on the easy grammar points (plural of nouns, simple past tense, modal verbs, many vs. much, comparative adjectives, since/for) and the difficult grammar points (indefinite article, 3rd person *–s* in the simple present tense, second conditional, yes/no questions, dative alternation, verb complements, relative clauses). The descriptive statistics are shown in Table 4.9.

Table 4.9. Descriptive statistics for the easy and difficult grammar points on the MLK test and its subsections

Easy grammar points	**N**	**Mean%**	**Max possible**	**Mean**	**SD**	**Min**	**Max**
MLK in total	90	75	24	18.09	3.47	11.5	23
MLK correction	90	89	6	5.35	.83	3	6
MLK description/explanation	90	63	12	7.53	2.55	2	11.5
MLK rule illustration	90	87	6	5.21	.71	3	6
Difficult grammar points							
MLK in total	90	61	28	16.95	3.98	7.5	24.5
MLK correction	90	84	7	5.86	1.13	2	7
MLK description/explanation	90	46	14	6.49	2.71	1	13
MLK rule illustration	90	65	7	4.57	1.18	1.5	7

The descriptive statistics in Table 4.9 show that the easy grammar points led to a higher mean percentage score (mean % = 75) than the difficult grammar points (mean % = 61). The results for the correction section indicate that there was not much difference between easy and difficult, but there was a clear difference for the rule illustration section, and to a slightly lesser extent for the description/explanation section. Figure 4.2 illustrates these differences.

Figure 4.2 Students' performance on the MLK test for easy and difficult grammar points



In order to check whether the difference was significant between the two levels of perceived difficulty, inferential statistics were calculated. As the data were normally distributed, the parametric paired-samples t-test was used, and the results indicate that there was a significant difference between learners' MLK total scores on easy and difficult grammar points $t(89) = -11.86$, $p < .001$, representing a medium effect, $d = .61$. This significant difference between the results of easy and difficult grammar points indicates that the difficulty judgements of grammar points according to learners and teachers were accurate with regard to explicit knowledge to the extent that learners' performance on easy grammar points was better than their performance on difficult grammar points.

Descriptive statistics were also calculated for each group of participants according to learners' performance on the MLK test on the easy and difficult grammar points. The descriptive statistics are shown in Table 4.10.

Table 4.10 Descriptive statistics of the performance of each group of participants on the easy and difficult grammar points on the MLK test

Level 5: Easy grammar points	N	Mean%	Max possible	Mean	SD	Min	Max
MLK in total	30	71	24	17.03	3.73	11.5	23
MLK correction	30	88	6	5.30	.82	3.5	6
MLK description/explanation	30	56	12	6.68	2.89	2	11.5
MLK rule illustration	30	84	6	5.05	.77	3	6
Level 5: Difficult grammar points							
MLK in total	30	53	28	14.83	3.98	7.5	20.5
MLK correction	30	79	7	5.53	1.34	2	7
MLK description/explanation	30	35	14	4.87	2.27	1	9
MLK rule illustration	30	62	7	4.35	1.22	1.5	6.5
Level 7: Easy grammar points							
MLK in total	30	78	24	18.77	2.87	13	22.5
MLK correction	30	90	6	5.42	.71	4	6
MLK description/explanation	30	66	12	7.95	2.09	3	11
MLK rule illustration	30	90	6	5.40	.58	4	6
Level 7: Difficult grammar points							
MLK in total	30	63	28	17.53	3.61	11	24.5
MLK correction	30	85	7	5.93	1.04	4	7
MLK description/explanation	30	51	14	7.08	2.75	2.5	12
MLK rule illustration	30	65	7	4.55	.99	2.5	6

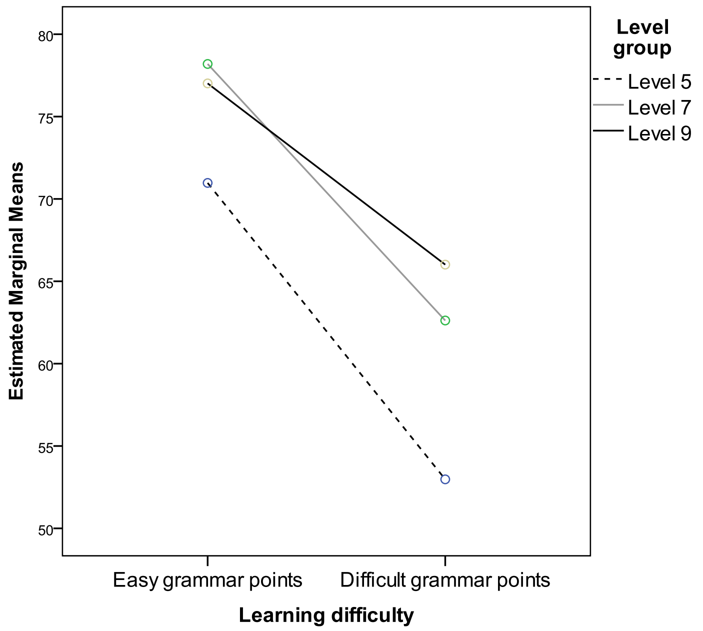
Level 9: Easy grammar points							
MLK in total	30	77	24	18.48	3.62	11.5	23
MLK correction	30	89	6	5.33	.97	3	6
MLK description/explanation	30	67	12	7.99	2.47	2	11
MLK rule illustration	30	86	6	5.18	.76	3.5	6
Level 9: Difficult grammar points							
MLK in total	30	66	28	18.48	3.49	11.5	24.5
MLK correction	30	87	7	6.12	.91	3.5	7
MLK description/explanation	30	54	14	7.53	2.36	2	13
MLK rule illustration	30	69	7	4.82	1.30	2	7

Table 4.10 indicates a very large difference for the MLK total mean percent scores between easy and difficult grammar points in Level 5 as well as between the MLK description/explanation and rule illustration section; the difference between the correction scores is smaller than for the other subsections. Similarly, there is a large difference for the MLK total mean percent scores between easy and difficult grammar points in Level 7 and Level 9 and the subsections except the correction section. It is worth noting that the difference between easy and difficult grammar points for the correction section in Level 5 decreases from five to two points in Level 7 and 9, respectively.

The variances within groups of participants show more homogeneity for easy grammar points in Level 5 and 7 than for difficult grammar points. It is the opposite in Level 9 where there is more homogeneity for difficult grammar points with the exception of the rule illustration section. The heterogeneity for difficult grammar points in this group of participants for each subsection of the MLK test suggests that these grammar points are difficult for some Level 9 learners but not so much for others.

A repeated-measures ANOVA with level groups and learning difficulty as independent variables was conducted (the assumption of normality of distribution of data and homogeneity of variance were met) to examine the effects of these two variables on learners' performance on the MLK test. The results show that there was a significant effect for learning difficulty F (1, 87) = 146.30, p < .01, r = .79, indicating that the learners' scores were significantly more accurate on easy grammar points than on difficult grammar points. Furthermore, the results show a significant effect for level group F (2, 87) = 5.253, p < .05, r = .33, as shown in the one-way ANOVA in section 4.1. The results also show a marginally significant interaction (p = .067) with a small effect size (r = .24) between level group and learning difficulty, as shown in Figure 4.3.

Figure 4.3 Interaction of learning difficulty and level group in the MLK test



There is a tendency for a more-to-less easy vs. difficult contrast from Level 5 to Level 9 as students get more experience with the L2 and more instruction.

4.4 RQ3a. What is the level of learners' implicit knowledge of the targeted grammar points as measured by the oral elicited imitation and oral narrative test?

Concerning RQ3a, descriptive statistics were calculated for both tests of implicit knowledge (i.e. the oral narrative test and the EI test). The descriptive statistics for the EI test are presented first.

Following the procedures of scoring and analysis in Erlam's (2006) study, the items on the EI test were grouped according to whether they tested participants' ability to repeat grammatically correct structures or the ability to correct ungrammatical structures. Table 4.11 shows the results for both grammatical and ungrammatical statements.

Table 4.11 Descriptive statistics for the performance on the EI test

	N	Mean%	Max Possible	Mean	SD	Min	Max
EI in total	90	61	93	56.72	13.42	23	80
Repeat grammatical	90	72	48	34.44	7.11	16	45
Correct ungrammatical	90	49	45	22.16	7.23	7	39

The results in Table 4.11 show that on average the whole cohort of participants found the EI test somewhat challenging (mean % = 61). The learners repeated 72 percent of grammatical items correctly. Twenty-eight percent of grammatical items were thus repeated incorrectly or no obligatory occasion was created for use of the target grammar point. The results also indicate that the learners did correct ungrammatical items. Forty-nine percent of ungrammatical items were corrected. Fifty-one percent of ungrammatical items were thus either repeated incorrectly or no obligatory occasion was created for use of the target grammar point.

A Pearson product moment correlation was conducted (the assumption of normal distribution was met, see Table 3.22) to assess shared variance between the grammatical and ungrammatical grammar points. A strong significant positive correlation (r = .85, p < .01) between these two categories of targeted grammar points was found. These results suggest that the same construct is being measured, allowing the use of a pooled total EI test score in all subsequent analyses.

To proceed with the statistical analysis in relation to the grouping variable, descriptive statistics were calculated for each group of participants. Table 4.12 shows students' performance on the EI test for each group tested.

Table 4.12 Descriptive statistics of the performance of each group of participants on the EI test

Level 5	**N**	**mean%**	**Max possible**	**Mean**	**SD**	**Min**	**Max**
EI in total	30	56	93	52.17	12.06	23	72
Level 7							
EI in total	30	60	93	56.17	14.18	28	78
Level 9							
EI in total	30	66	93	61.47	12.32	37	80

The descriptive statistics in Table 4.12 show that all groups of participants (mean % = 56 for Level 5, mean % = 60 for Level 7, and mean % = 66 for Level 9) found the EI test relatively difficult. It is worth noting that the higher the level group, the higher the score was, as expected given the difference of amount of time of practice in the L2 and the difference in overall L2 proficiency.

A one-way ANOVA test was conducted to find out whether the overall mean score differences between the groups of participants were significant; the assumptions of normal distribution of data and homogeneity of variance were met. The results indicate that there was a significant difference in learners' performance (F = 3.93, p < .05, representing a small effect, d = .04). *Post hoc* tests (Tukey's test) to compare all groups of participants with each other were conducted, and these results show that there was a significant difference between participants of Level 5 and Level 9 (mean difference = -9.3, 95% CI = -17.24, -1.36, p < .05). No significant differences were found between participants of Level 5 and Level 7 (mean difference = -4.0, 95% CI = -

11.94, 3.9, *p* = .46) nor between Level 7 and 9 (mean difference = -5.3, 95% CI = -13.24, 2.6, *p* = .25). This indicates that only Level 9 outperformed Level 5.

Turning to the second measure of implicit knowledge, the oral narrative test, Table 4.13 shows the descriptive statistics for the participants' scores; recall that a minimum of 3 occurrences was used as the threshold for giving a score for each grammar point. Also recall that a subset of 8 grammar points (simple past tense, 3rd person –*s* in the simple present tense, comparative adjectives, verb complements, indefinite article, modal verbs, plural of nouns, dative alternation) of the 13 targeted grammar points were included in the oral narrative test. The grammar points modal verbs, comparative adjectives, and dative alternation were excluded from the analysis because more than 50% of students did not supply 3 occurrences as required. Therefore, only 5 grammar points are represented in the analysis of the oral narrative scores: two easy grammar points (simple past tense and plural of nouns) and three difficult grammar points (3rd person –*s* in the simple present tense, verb complements, and indefinite article). It is worth noting that not all learners were successful in producing the minimum required number of suppliances for every targeted grammar point, that is, the total number of students that supplied a minimum of 3 occurrences varied between grammar points.

Table 4.13 Descriptive statistics for the performance on the oral narrative test

	N	**Mean%**	**Max Possible**	**SD**	**Min**	**Max**
Accuracy in total	89*	72	100%	17.41	23	100

*One participant failed to produce the minimum required number of suppliances for any of the grammar points and was thus excluded from the mean overall accuracy score.

As can be seen in Table 4.13, learners performed higher on this test (mean % = 72) than on the EI test (mean % = 61). The result reflects that learners' production of sentences varies in each type of test when they are subjected to different ways of speech elicitation. In addition, learners may have found the oral narrative test easier because fewer grammar points were involved, and they had a choice to select from their linguistic repertoire the structures they needed to retell the story. Surprisingly, the increasing development in the L2 found between groups on the EI test was not found in the case of learners' performance on the ON test as displayed in Table 4.14.

Table 4.14 Descriptive statistics of the performance of each group of participants on the ON test

Level 5	N	Mean%	Max possible	SD	Min	Max
ON in total	30	66	100	15.44	36	91
Level 7						
ON in total	30	76	100	15.94	33	96
Level 9						
ON in total	29	73	100	19.53	23	100

The difference in scores between Level 5 and Level 7 is as expected, but the learners' better performance in Level 7 than Level 9 is unexpected. Admittedly, learners in Level 9 varied more in their responses on this test than learners in Level 7 as illustrated by their variance and their minimum and maximum scores.

A one-way ANOVA was conducted to find out whether the overall mean score differences between the groups of participants were significant; the assumptions of normal distribution of data and homogeneity of variance were met. The results indicate that there was a marginal difference in learners' performance (F = 2.910, p = .060). *Post hoc* tests (Gabriel's test[6]) show that there was a marginal difference between Level 5 and 7 (mean difference = -10.242, 95% CI

[6] Gabriel's test is the post-hoc test appropriate for slightly different sample sizes and homogeneous variances (Larson-Hall, 2010).

= -20.95, .46, p = .065). No significant differences were found between Level 5 and 9 nor between Level 7 and 9.

So far, descriptive and inferential statistics have been run on each implicit measure separately. Combining the scores of both tests would yield a single more powerful implicit score. In order to do this, a Pearson product moment correlation was conducted between the overall EI score and the overall ON score (the assumption of normal distribution was met, see Table 3.22). Results reveal a significant positive correlation of medium strength ($r = .47$, $p < .01$). Given this positive correlation, both tests scores (i.e. the EI and ON scores) were combined into one by adding the overall EI and ON percentage scores and dividing them by 2. This was done based on the assumption that the combined scores of the targeted grammar points provide a fuller representation of learners' implicit knowledge. Table 4.15 shows the combined scores in percentage of implicit knowledge for all participants.

Table 4.15 Descriptive statistics for combined implicit knowledge scores

	N	**Mean%**	**Max Possible**	**SD**	**Min**	**Max**
EI/ON combined	90	66	100%	14.27	25	93

As can be seen in Table 4.15, the mean percent implicit score (66) is very similar to the mean percent explicit score (67, i.e. the MLK score). It should be acknowledged that the score for explicit knowledge partly depends on the higher score in the correction section, in which learners may have employed both their implicit and explicit knowledge of each grammar point, and it should also be acknowledged that the score for implicit knowledge may partly depend on the use of explicit knowledge of each grammar point on the part of the learners.

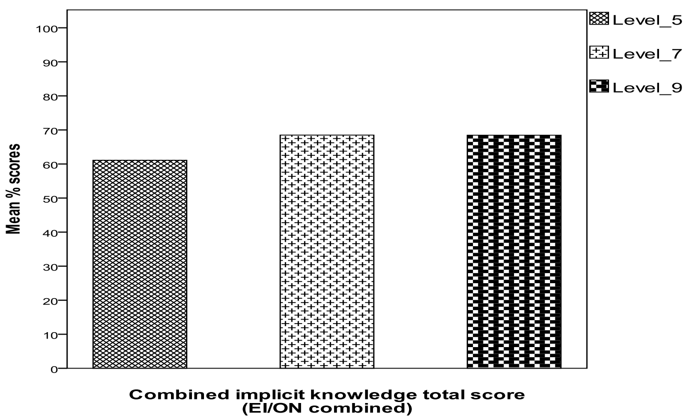
Descriptive statistics were calculated in accordance with the combined implicit knowledge scores for each group of participants. Table 4.16 shows the combined implicit knowledge scores for each of the groups tested.

Table 4.16 Descriptive statistics of each group of participants for combined implicit knowledge scores

Level 5	N	Mean%	Max possible	SD	Min	Max
Combined implicit in total	30	61	100	10.66	44	84
Level 7						
Combined implicit in total	30	68	100	13.91	37	89
Level 9						
Combined implicit in total	30	69	100	16.70	25	93

Like the difference in scores between Level 5 and 7 on the ON test, the difference for the combined implicit score between these two levels is seven percent points. In contrast, the difference of the combined implicit knowledge scores between Level 7 and 9 is only one point. The variance for Level 9 is also greater than in the other two levels; this is most likely attributable to the learners' heterogeneous performance on the ON test. Figure 4.4 shows the students' combined implicit knowledge scores for the three groups of participants.

Figure 4.4 Students' combined implicit knowledge scores



A one-way ANOVA was conducted to find out whether the overall mean score differences between the groups of participants were significant; the assumptions of normal distribution of data and homogeneity of variance were met. The results indicate that there was a marginal difference in learners' performance (F = 2.869, p = .062), but *post hoc* tests (Tukey's test) did not identify significant differences between the individual levels. This may be attributable to the pattern of results obtained for the EI and ON tests separately, with differences between Level 5 and Level 9 (EI test) on the one hand and between Level 5 and Level 7 (ON test) on the other hand being levelled out.

4.5 RQ3b. What is the level of learners' implicit knowledge of the easy and difficult grammar points as measured by the oral elicited imitation and the oral narrative test?

In response to RQ3b, descriptive statistics were calculated for the two categories (i.e. easy and difficult grammar points) of perceived difficulty in accordance with learners' performance on the EI and ON test. The descriptive statistics are shown in Table 4.17.

Table 4.17 Descriptive statistics for performance on the easy and difficult grammar points on the EI and ON test.

EI test	N	Mean% possible	Max	Mean	SD	Min	Max
Easy grammar points	90	64	43	27.46	6.30	13	39
Difficult grammar points	90	58	50	29.21	7.59	9	45
ON test							
Easy grammar points	90	79	100	79	17.70	25	100
Difficult grammar points	90	66	100	66	24.41	4	100

The mean percent scores in Table 4.17 show a difference between the easy and difficult grammar points on both the EI and the ON test; on the easy grammar points, the participants scored higher than on the difficult ones. It can also be seen that the participants' responses for the easy grammar points were more homogeneous as indicated by the lower SD and range. A paired-samples *t*-test was run to see whether the difference between the easy and difficult grammar points of the EI and ON test for the whole cohort of participants was significant (the assumption of normality and homogeneity of variance were met). The results for the easy and difficult grammar points of the EI test indicated that there was a significant difference $t(89) = -4.19$, $p < .001$, representing a small effect, $d = .28$. The same statistical test was run for the ON test and a significant difference was found between the easy and difficult grammar points $t(87) = 4.85$, $p < .001$, representing a medium effect, $d = .72$.

Taking into consideration that the EI test was the only implicit measure to include all targeted grammar points, descriptive statistics were also calculated for learners' performance on the EI test for each group of participants for the easy and difficult grammar points as shown in Table 4.18.

Table 4.18 Descriptive statistics of each group of participants for the EI scores for the easy and difficult grammar points

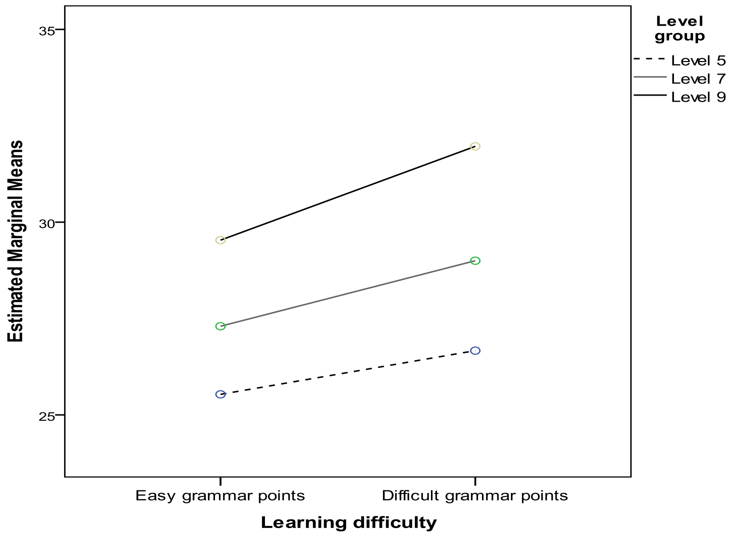
Level 5	**N**	**Mean%**	**Max possible**	**Mean**	**SD**	**Min**	**Max**
Easy grammar points							
EI in total	30	60	43	26	5.67	14	36
Difficult grammar points							
EI in total	30	54	50	27	6.74	9	38
Level 7							
Easy grammar points							
EI in total	30	63	43	27	6.66	13	37
Difficult grammar points							
EI in total	30	58	50	29	8.39	13	45
Level 9							
Easy grammar points							

EI in total	30	70	43	30	6.07	17	39
Difficult grammar points							
EI in total	30	64	50	32	6.81	19	43

In each level group, learners obtained higher scores on the easy grammar points than on the difficult grammar points. There is a difference in the variance between easy and difficult grammar points in each level; the variance is larger in Level 7 than in the other two levels, and in Level 9, easy grammar points show a slightly larger variance than difficult grammar points.

A repeated-measures ANOVA with level groups and learning difficulty as independent variables was conducted (the assumptions of normality of distribution of data and homogeneity of variance were met) to examine the effects of these two variables on learners' EI test scores. The results show that there was a significant effect for learning difficulty F (1, 87) = 17.43, p < .01, r = .41, indicating that the learners' scores were significantly more accurate on easy grammar points than on difficult grammar points. Furthermore, the results also show a significant effect for level group F (2, 87) = 3.883, p = .024, r = .28, indicating that the learners' scores showed greater accuracy at higher levels. The results also show a non-significant interaction (p = .451) between level group and learning difficulty as displayed in Figure 4.5.

Figure 4.5 Interaction of learning difficulty and level group for the EI test scores



As an additional check, descriptive statistics were also calculated for the whole sample as well as for each group of participants for the easy and difficult grammar points in the combination of scores learners obtained in both measures of implicit knowledge as shown in Table 4.19 and 4.20, respectively.

Table 4.19 Descriptive statistics for the combination of scores of implicit knowledge for the easy and difficult grammar points.

	N	Mean%	Max possible	Mean	SD	Min	Max
Easy grammar points							
Implicit knowledge in total	90	70	100	70	14.46	22.1	95.3
Difficult grammar points							
Implicit knowledge in total	90	62	100	62	17.24	20	95

Like the results obtained on the EI and the ON test separately, learners performed better on the easy grammar points than on the difficult grammar points. A *t*-test was conducted to see whether there was a significant difference between easy and difficult grammar points (the assumption of normality and homogeneity of variance were met). The results showed a significant difference between the easy and difficult grammar points $t(89) = 5.64$, $p < .001$, representing a small effect, $d = .26$. This significant difference between the results of easy and difficult grammar points confirms the learners' and teachers' difficulty judgements of the targeted grammar points; with regard to implicit knowledge, learners performed better on easy grammar points than on difficult grammar points.

The descriptive statistics for each group of participants for the easy and difficult grammar points in the combination of scores learners obtained in both measures of implicit knowledge are displayed in Table 4.20.

Table 4.20 Descriptive statistics of each group of participants for the combination of scores of implicit knowledge for the easy and difficult grammar points

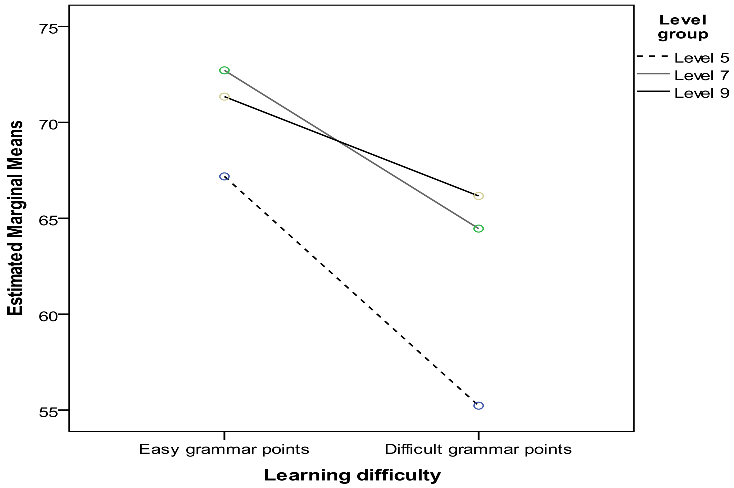
Level 5	N	Mean%	Max possible	Mean	SD	Min	Max
Easy grammar points							
Implicit knowledge in total	30	67	100	n/a	12.20	36.90	87.40
Difficult grammar points							
Implicit knowledge in total	30	55	100	n/a	14.96	29.70	81.80
Level 7							
Easy grammar points							
Implicit knowledge in total	30	73	100	n/a	11.39	47.10	90.70
Difficult grammar points							
Implicit knowledge in total	30	64	100	n/a	17.68	27.30	95
Level 9							
Easy grammar points							
Implicit knowledge in total	30	71	100	n/a	18.61	22.10	95.30
Difficult grammar points							
Implicit knowledge in total	30	66	100	n/a	17.46	20	94

In each level group, learners obtained higher scores on the easy grammar points than on the difficult grammar points. There is a difference in the variance between easy and difficult grammar points in each level; the variance is larger in Level 7 than in the other two levels, and in Level 9, easy grammar points show a slightly larger variance than difficult grammar points.

A repeated-measures ANOVA with level groups and learning difficulty as independent variables was conducted (the assumptions of normality of distribution of data and homogeneity of variance were met) to examine the effects of these two variables on learners' implicit knowledge scores.

The results show that there was a significant effect for learning difficulty $F(1, 87) = 32.360$, $p < .01$, $r = .52$, indicating that the learners' scores were significantly more accurate on easy grammar points than on difficult grammar points. Furthermore, the results show a marginal effect for level group $F(2, 87) = 2.866$, $p = .062$, $r = .25$, indicating that the learners' scores showed a trend towards greater accuracy at higher levels. The results also show a non-significant interaction ($p = .183$) between level group and learning difficulty. Figure 4.6 illustrates these results.

Figure 4.6 Interaction of learning difficulty and level group for the combined implicit scores



4.6 RQ4. What is the relationship between learners' implicit and explicit knowledge of the 13 targeted grammar points?

As a first step toward answering RQ4, Pearson correlations were calculated for the EI test, the ON test (implicit knowledge), and the MLK test scores (explicit knowledge). The assumption of normal distribution was met for all variables except for the MLK correction section (see Table 3.22), but taking into account that Pearson product moment correlation is a robust measure that can cope with some violations (see Pallant, 2010; Norman, 2010; see also Trafimov & Marks, 2015), the correlation was conducted. Table 4.21 shows the correlations between the scores of the implicit measures and the scores of the explicit measure.

Table 4.21 Correlations (Pearson *r*): Tests of MLK, EI, and ON

	ON test	**MLK test**	**Correction**	**Description/ explanation**	**Rule illustration**
EI test	.47** *p* = .00	.31** *p* = .00	.33** *p* = .00	.28** *p* = .01	.14 *p* = .20
ON test		.16 *p* = .15	.21 *p* = .05	.13 *p* = .21	.03 *p* = .79
MLK test			.75** *p* = .00	.96** *p* = .00	.66** *p* = .00
Correction				.62** *p* = .00	.31** *p* =.00
Description/ explanation					.51** *p* = .00

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

As discussed in the previous chapter, the MLK test correlated strongly with each of its subsections (r = .66 to .96). Table 4.21 shows a significant moderate association between the EI test and the ON test. Table 4.21 also displays non-significant correlations between the scores of the ON test and the scores of the MLK test, which is probably due to the small number of grammar points included on the ON test. However, it is important to note that the association between the ON test and correction section of the MLK test approaches significance (p = .05). On the other hand, there are significant positive correlations between the scores of the EI test and the scores of the MLK test and its subsections with the exception of the rule illustration section.

Further Pearson product moment correlations were run (the assumption of normality was met for most variables, see Table 3.24) for each group of participants in order to identify any correlation between learners' performance on the EI test and the MLK test at group level. No significant correlations were found between the EI test and MLK test at group level.

In order to compare learners' EI test scores with their MLK test scores for the 13 grammar points individually, Table 4.22 shows learners' EI test scores and MLK test scores for each of the targeted grammar points in descending order.

Table 4.22 Descriptive statistics for learners' EI test scores and MLK test scores for the 13 grammar points

EI test scores	Mean%	SD	MLK test scores	Mean%	SD
1 Relative clauses	84%	9	1 Yes/No questions	81%	18
2 Indefinite article	76%	23	2 Modal verbs	81%	25
3 Dative alternation	74%	13	3 Comparative adjectives	78%	8
4 Plural of nouns	74%	52	4 Many vs. Much	76%	11
5 Modal verbs	65%	19	5 Simple present tense	76%	16
6 Comparative adjectives	63%	35	6 Since/for	75%	9
7 Simple past tense	61%	51	7 Plural of nouns	73%	13
8 Many vs. Much	60%	29	8 Simple past tense	69%	32
9 Since/for	53%	29	9 Indefinite article	60%	17
10 Second conditional	52%	15	10 Verb complements	58%	11
11 Yes/No questions	48%	31	11 Second conditional	53%	9
12 Simple present tense	47%	32	12 Relative clauses	51%	44
13 Verb complements	40%	26	13 Dative alternation	45%	28

According to the mean and SD percentages of the grammar points in Table 4.22, the grammar points plural of nouns and second conditional were the only grammar point learners found equally difficult on the EI test and MLK test. The rest of the grammar points are ranked differently on each type of test, and most EI test scores show a greater variance than the MLK test scores. A Spearman's rank order correlation between EI test and MLK test scores by grammar point showed a non-significant negative correlation (rho = -.32, p = .28).

The same comparison, that is, learners' EI test scores with their MLK test scores of the 13 individual grammar points, was made for each group of participants. Table 4.23 shows the scores for each type of test of the targeted grammar points by level group.

Table 4.23 Mean % scores and SDs for learners' EI test and MLK test scores for the 13 grammar points by level group

Level 5					
EI test scores	**Mean%**	**SD**	**MLK test scores**	**Mean%**	**SD**
1 Relative clauses	77	14	1 Yes/no questions	75	17
2 Indefinite article	72	26	2 Since/for	75	11
3 Dative alternation	69	16	3 Plural of nouns	73	12
4 Plural of nouns	68	48	4 Simple present tense	73	13
5 Comparative adjectives	66	44	5 Comparative adjectives	73	5
6 Modal verbs	57	25	6 Modal verbs	74	18
7 Many vs. much	57	35	7 Simple past tense	66	25
8 Since/for	53	31	8 Many vs. much	66	5
9 Simple past tense	50	39	9 Verb complements	55	6
10 Second conditional	46	23	10 Indefinite article	49	27
11 Yes/no questions	44	34	11 Second conditional	49	15
12 Simple present tense	42	32	12 Relative clauses	41	39
13 Verb complements	33	32	13 Dative alternation	35	26
Level 7					
1 Relative clauses	88	8	1 Modal verbs	89	36
2 Plural of nouns	77	53	2 Yes/no questions	80	20
3 Indefinite article	76	23	3 Comparative adjectives	80	9
4 Dative alternation	73	14	4 Many vs. much	79	15
5 Modal verbs	69	15	5 Since/for	76	9

6 Many vs much	62	24	6 Plural of nouns	73	10
7 Simple past tense	60	55	7 Simple present tense	73	13
8 Comparative adjectives	58	29	8 Simple past tense	73	35
9 Second conditional	51	15	9 Indefinite article	63	26
10 Since/for	46	27	10 Second conditional	62	15
11 Yes/no questions	46	29	11 Verb complements	60	15
12 Simple present tense	44	37	12 Relative clauses	55	53
13 Verb complements	41	25	13 Dative alternation	51	39
Level 9					
1 Relative clauses	85	12	1 Yes/no questions	85	20
2 Dative alternation	80	11	2 Many vs. much	84	18
3 Indefinite article	80	21	3 Simple present tense	83	26
4 Plural of nouns	77	56	4 Comparative adjectives	83	12
5 Simple past tense	73	61	5 Modal verbs	78	22
6 Modal verbs	68	20	6 Since/for	75	10
7 Comparative adjectives	65	34	7 Plural of nouns	71	19
8 Many vs. much	63	31	8 Simple past tense	71	39
9 Since/for	59	32	9 Indefinite article	68	22
10 Second conditional	58	13	10 Second conditional	63	15
11 Simple present tense	55	31	11 Verb complements	59	12
12 Yes/no questions	53	30	12 Relative clauses	57	4
13 Verb complements	46	24	13 Dative alternation	49	28

Spearman's rank order correlations were conducted between EI and MLK test scores by grammar point for each group of participants. The results showed non-significant negative correlations for Level 5 (rho = -38, *p* = .197), Level 7 (rho = -.21, *p* = .475), and Level 9 (rho = -.45, *p* = .117). These correlations show the same direction as the correlation for the whole sample of participants between EI test scores and MLK test scores of individual grammar points. It is worth noting that the grammar point relative clauses takes the highest position on the EI test in the three different group levels and the second lowest position on the MLK test. It is also worth noting that the grammar point verb complements takes the lowest position on the EI test in the three different groups while the grammar point dative alternation takes the same position on the MLK test.

As a second step toward answering RQ4, Pearson correlations were also calculated for the overall EI/ON combined scores (implicit knowledge) and the MLK test scores (explicit knowledge). Table 4.24 shows the correlations between scores of these variables.

Table 4.24 Correlations (Pearson *r*): Tests of MLK and EI/ON combined

	MLK test	**Correction**	**Description/ explanation**	**Rule illustration**
EI/ON combined	.22* *p* = .04	.28** *p* = .01	.19 *p* = .07	.05 *p* = .66
MLK test		.75** *p* = .00	.96** *p* = .00	.66** *p* = .00
Correction			.62** *p* = .00	.31** *p* =.00
Description/ explanation				.51** *p* = .00

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Scores on the metalinguistic knowledge test as a whole and the EI/ON combined scores are associated (r = .22). This rather weak association can be attributed to the fact that all measures targeted the same grammar points. It is, however, primarily driven by the correction section, with the correlation for the description/explanation section approaching significance. The association between the EI/ON scores and the MLK rule illustration section scores is not significant. The significant positive correlations between the scores of the EI test and the scores of the MLK test and its subsections with the exception of the rule illustration section as shown in Table 4.21 means that these correlations drive the significant correlations between the implicit and explicit scores.

Further Pearson product moment correlations were run (the assumption of normality was met for most variables, see Table 3.24) for each group of participants in order to identify any correlation between the measures of implicit and explicit knowledge at group level. No significant correlations were found between the implicit and explicit knowledge of targeted grammar points in any of the groups of participants, except for one significant correlation between MLK correction section and EI/ON combined (r = .40, p = .03) in Level 9.

In order to compare learners' explicit knowledge with their implicit knowledge for the 13 grammar points individually, Table 4.25 shows learners' implicit (i.e. EI/ON combined) and explicit knowledge (i.e. MLK test) scores for each of the targeted grammar points in descending order.

Table 4.25 Descriptive statistics for learners' implicit and explicit knowledge scores for the 13 grammar points

Implicit scores	**Mean%**	**SD**	**Explicit score**	**Mean%**	**SD**
1 Plural of nouns	84%	32	1 Yes/No questions	81%	18
2 Relative clauses	84%	9	2 Modal verbs	81%	25
3 Indefinite article	83%	20	3 Comparative adjectives	78%	8
4 Dative alternation	74%	13	4 Many vs. Much	76%	11
5 Simple past tense	66%	39	5 Simple present tense	76%	16
6 Verb complements	65%	22	6 Since/for	75%	9

7 Modal verbs	65%	19	7 Plural of nouns	73%	13
8 Comparative adjectives	63%	35	8 Simple past tense	69%	32
9 Many vs. Much	60%	29	9 Indefinite article	60%	17
10 Since/for	53%	29	10 Verb complements	58%	11
11 Second conditional	52%	15	11 Second conditional	53%	9
12 Simple present tense	49%	31	12 Relative clauses	51%	44
13 Yes/No questions	48%	31	13 Dative alternation	45%	28

Table 4.25 shows that the grammar point second conditional was the only grammar point learners found equally difficult as implicit and explicit knowledge. The rest of the grammar points are ranked differently for each type of knowledge, and most implicit scores show a greater variance than the explicit scores. A Spearman's rank order correlation between implicit and explicit knowledge scores by grammar point showed a negative correlation approaching significance (rho = -.54, p = .058). The trend towards a negative association suggests that when learners had strong explicit knowledge of a grammar point they showed a tendency towards having weak implicit knowledge of the same grammar point, and vice versa. For instance, whereas learners' implicit knowledge results for the grammar points plural of nouns, relative clauses, indefinite article, and dative alternation were high, the results for these grammar points on the explicit measure were low. Likewise, the results for the grammar points yes/no questions, modal verbs, comparative adjectives, and many vs. much were high on the explicit measure but low on the implicit measure.

The same comparison, that is, learners' explicit knowledge with their implicit knowledge of the 13 individual grammar points, was made for each group of participants. Table 4.26 shows learners' implicit (i.e. EI/ON combined) and explicit knowledge (i.e. MLK test) scores for each of the targeted grammar points by level group.

Table 4.26 Mean % scores and SDs for learners' implicit and explicit knowledge for the 13 grammar points by level group

Level 5					
Implicit score	**Mean%**	**SD**	**Explicit score**	**Mean%**	**SD**
1 Indefinite article	80	22	**1** Yes/no questions	75	17
2 Plural of nouns	79	31	**2** Since/for	75	11
3 Relative clauses	77	14	**3** Plural of nouns	73	12
4 Dative alternation	69	16	**4** Simple present tense	73	13
5 Comparative adjectives	66	44	**5** Comparative adjectives	73	5
6 Simple past tense	58	33	**6** Modal verbs	74	18
7 Verb complements	58	25	**7** Simple past tense	66	25
8 Modal verbs	57	25	**8** Many vs. much	66	5
9 Many vs. much	57	35	**9** Verb complements	55	6
10 Since/for	53	31	**10** Indefinite article	49	27
11 Second conditional	46	23	**11** Second conditional	49	15
12 Yes/no questions	44	34	**12** Relative clauses	41	39
13 Simple present tense	41	31	**13** Dative alternation	35	26
Level 7					
1 Relative clauses	88	8	**1** Modal verbs	89	36
2 Plural of nouns	87	30	2 Yes/no questions	80	20
3 Indefinite article	83	19	**3** Comparative adjectives	80	9
4 Dative alternation	73	14	**4** Many vs. much	79	15
5 Modal verbs	69	15	**5** Since/for	76	9
6 Verb complements	67	19	**6** Plural of nouns	73	10
7 Simple past tense	66	40	**7** Simple present tense	73	13
8 Many vs much	62	24	**8** Simple past tense	73	35

9 Comparative adjectives	58	29	**9** Indefinite article	63	26
10 Second conditional	51	15	**10** Second conditional	62	15
11 Simple present tense	49	31	**11** Verb complements	60	15
12 Since/for	46	27	**12** Relative clauses	55	53
13 Yes/no questions	46	29	**13** Dative alternation	51	39
Level 9					
1 Indefinite article	86	17	**1** Yes/no questions	85	20
2 Plural of nouns	85	30	**2** Many vs. much	84	18
3 Relative clauses	85	12	**3** Simple present tense	83	26
4 Dative alternation	80	11	**4** Comparative adjectives	83	12
5 Simple past tense	75	42	**5** Modal verbs	78	22
6 Verb complements	70	21	**6** Since/for	75	10
7 Modal verbs	68	20	**7** Plural of nouns	71	19
8 Comparative adjectives	65	34	**8** Simple past tense	71	39
9 Many vs. much	63	31	**9** Indefinite article	68	22
10 Since/for	59	32	**10** Second conditional	63	15
11 Second conditional	58	13	**11** Verb complements	59	12
12 Simple present tense	56	31	**12** Relative clauses	57	4
13 Yes/no questions	53	30	**13** Dative alternation	49	28

Spearman's rank order correlations were conducted between implicit and explicit knowledge scores by grammar point for each group of participants. The results showed non-significant negative correlations for Level 5 (rho = -.43, p = .147) and Level 7 (rho = -.42, p = .155), and a significant negative correlation for Level 9 (rho = -.61, p = .027). These correlations show the same direction as the correlation for the whole sample of participants between implicit and explicit knowledge of individual grammar points. It is worth noting that only the correlation for Level 9 reaches significance, however. Thus, it is in the highest level group where explicit and implicit knowledge of individual grammar points diverge significantly.

4.7 RQ5. What is the relationship between learners' implicit and explicit knowledge of the 13 grammar points, their language learning aptitude and working memory capacity?

As a first step towards answering RQ5, descriptive statistics were calculated for each of the LLAMA subtests and for the test as a whole (i.e. language aptitude) as well as for the BDS test (i.e. working memory). Table 4.27 shows the descriptive statistics for the whole sample of participants.

Table 4.27 Descriptive statistics for the LLAMA test (whole sample; *N* = 90) and the BDS test

	Mean%	Max possible	Mean	SD	Min	Max
LLAMA total	50	375	188.22	63.29	45	340
LLAMA_B (Vocabulary learning)	48	100	47.78	20.85	10	100
LLAMA_D (Sound recognition)	40	75	30.11	12.54	0	60
LLAMA_E (Sound-symbol correspondence)	69	100	68.72	28.85	0	100
LLAMA_F (Grammatical inferencing)	42	100	41.61	24.12	0	100
BDS total	41	28	11.57	4.05	5	24

The descriptive statistics in Table 4.27 show that the learners obtained a mean facility value of 50% on the LLAMA test as a whole. With respect to the subtest scores, learners scored highest on the LLAMA_E subtest (Sound-symbol correspondence; mean % = 69) and lowest on LLAMA_D subtest (Sound recognition; mean % = 40). On the LLAMA_E learners had to identify the written form in an unfamiliar alphabet of syllables they heard while on the LLAMA_D they had to identify a word they had previously heard in an unfamiliar language.

The results also show that participants found the LLAMA_D task (Sound recognition; mean % = 40) the most difficult followed by the LLAMA_F task (Grammatical inferencing; mean % = 42) and LLAMA_B (Vocabulary learning; mean % = 48). Table 4.27 also shows that learners obtained a mean facility value of 41 percent on the BDS test.

According to the LLAMA manual (Meara, 2005), the ranges of percent scores between 25 to 45 for the LLAMA_B, 15 to 35 for LLAMA_D, and 20 to 45 for LLAMA_E and LLAMA_F represent an average score. Participants' scores are in these ranges for the subtests LLAMA_B, LLAMA_D and LLAMA_F except for the LLAMA_E; this score is in a higher range, which is a good score. Taking into account the individual LLAMA subtest scores, the LLAMA total scores (mean % = 50) can be interpreted as an average score.

In order to see whether the LLAMA subtests and the BDS test correlated with each other, a Pearson product moment correlation was run between them (the assumption of normality was met for all subtests except for LLAMA_E, see Table 3.22). As argued above, Pearson product moment correlation is a robust measure that can cope with some violations (see Pallant, 2010; Norman, 2010; see also Trafimov & Marks, 2015), hence this type of analysis was carried out. Table 4.28 displays these correlations.

Table 4.28 Correlations (Pearson *r*): Subtests of LLAMA

	LLAMA_B	**LLAMA_D**	**LLAMA_E**	**LLAMA_F**	**BDS test**
LLAMA total	.70** *p* = .00	.49** *p* = .00	.80** *p* = .00	.81** *p* = .00	.33** *p* = .00
LLAMA_B		.26** *p* = .01	.34** *p* = .00	.44** *p* = .00	.29** *p* = .00
LLAMA_D			.22* *p* = .04	.29* *p* = .01	.13 *p* = .24
LLAMA_E				.49** *p* = .00	.27* *p* = .01
LLAMA_F					.23* *p* = .03

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Table 4.28 shows that the subtests LLAMA_B, LLAMA_E, and LLAMA_F correlated more strongly with each other than with LLAMA_D. A significant positive correlation was found between language aptitude and working memory (r = .33, p = .001). Working memory correlated with the LLAMA subtests except LLAMA_D (r = .13, p = .24). The significant positive correlation between the two IDs (language aptitude and working memory) is to be expected given the nature of each measure. On the language aptitude measure, as on the working memory measure, learners had to store and manipulate information to complete each task.

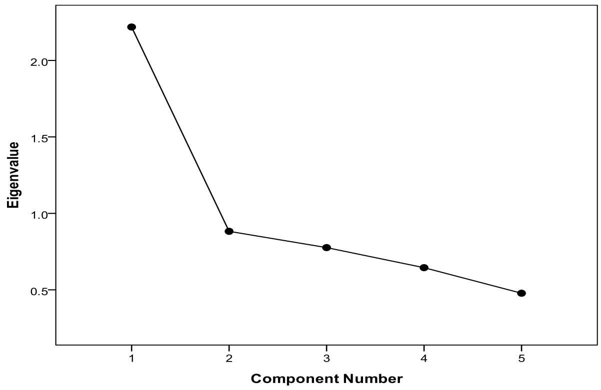
To better understand the structure of the variables of aptitude and working memory as measured in the present sample, a principal components analysis was conducted. The Kaiser-Meyer-Olkin measure of sampling adequacy was checked and its value was .737, which was higher than the recommended minimum value of .6 (see Pallant, 2010). In addition, Barlett's test of sphericity was checked as well and it was significant (p < .001). Five variables were included in the analysis: the four LLAMA subtests and the BDS score. An Oblimin rotation principal components analysis resulted in one component with an eigenvalue greater than 1 (2.219). This component accounted for 44.37% of the total variance. Table 4.29 shows the factor loadings.

Table 4.29 Loadings for principal components analysis

Test	**Component 1**
LLAMA_F	.776
LLAMA_E	.726
LLAMA_B	.725
BDS	.538
LLAMA_D	.523

Figure 4.7 illustrates various components. The scree plot clearly indicates that there is a break after the first component.

Figure 4.7 Components of factor analysis



The results of the present study are different from Granena's (2013) study in which the LLAMA subtests loaded on two different factors, that is, LLAMA_B, LLAMA_E, and LLAMA_F loaded on the first component while LLAMA_D loaded on the second component. Given that this principal components analysis was exploratory and in keeping with convention in existing research, in the present study the constructs of language aptitude and working memory were still treated as distinct, in the sense that scores from the aptitude and working memory measures were not combined into a single score (Hummel, 2009; Roehr & Gánem-Gutiérrez, 2009).

To find out whether learners at different levels differed in their language learning aptitude and working memory, descriptive statistics were calculated for each group of participants. Table 4.30 shows learners' performance on the LLAMA test and its subtests and the BDS test for each group of participants.

Table 4.30 Descriptive statistics for performance of participants on the LLAMA test and its subtests for each group of participants

	N	Mean%	Max possible	Mean	SD	Min	Max
Level 5							
LLAMA total	30	46	375	173.83	70.50	50	335
LLAMA_B (Vocabulary learning)	30	41	100	41.17	20.45	10	90
LLAMA_D (Sound recognition)	30	38	75	28.50	12.61	5	55
LLAMA_E (Sound-symbol correspondence)	30	64	100	64.00	31.25	0	100

LLAMA_F (Grammatical inferencing)	30	40	100	40.17	24.79	0	90
BDS total	30	40	28	11.17	4.46	5	23
Level 7							
LLAMA total	30	54	375	202.67	56.23	95	340
LLAMA_B (Vocabulary learning)	30	50	100	49.50	23.06	10	100
LLAMA_D (Sound recognition)	30	41	75	30.50	12.69	0	45
LLAMA_E (Sound-symbol correspondence)	30	75	100	75.00	23.89	10	100
LLAMA_F (Grammatical inferencing)	30	48	100	47.67	24.31	0	100
BDS total	30	41	28	11.57	3.72	5	22
Level 9							
LLAMA total	30	50	375	188.17	61.08	45	320
LLAMA_B (Vocabulary learning)	30	53	100	52.67	17.65	25	100
LLAMA_D (Sound recognition)	30	42	75	31.33	12.59	10	60
LLAMA_E (Sound-symbol correspondence)	30	67	100	67.17	30.67	0	100
LLAMA_F (Grammatical inferencing)	30	37	100	37.00	22.77	0	80
BDS total	30	43	28	11.97	4.05	6	24

Table 4.30 indicates that all groups of participants performed similarly on the language aptitude test and the working memory test. It is worth noting that learners' performance on the LLAMA_D subtest is quite consistent across levels. Another aspect to notice is that the scores for the LLAMA_E subtest was much higher than the scores for the other subtests. Like learners' performance on the LLAMA_D subtest, learners' performance on the BDS test is quite consistent across levels.

A series of one-way ANOVAs were conducted for the LLAMA test, the LLAMA subtests, and the BDS test. A one-way ANOVA test was run to find out whether the overall mean score differences of the BDS test between the groups of participants were significant; the assumptions of normal distribution of data and homogeneity of variance were met. The results show that there was no significant difference in learners' performance (*F* = .288, *p* = .751).

Another one-way ANOVA was conducted to find out whether the overall mean score difference of the LLAMA test between the groups of participants were significant; the assumptions of normal distribution of data and homogeneity of variance were met. The results indicate that there was no significant difference in learners' performance (*F* = 1.577, *p* = .212).

One more one-way ANOVA was run to find out whether the overall mean score difference of the LLAMA subtests between the groups of participants were significant; the assumptions of the normal distribution of data and homogeneity of variance were met. The results indicate that there were no significant differences in learners' performance for any of the subtests (LLAMA_B, *F* = 2.517, *p* = .087; LLAMA_D, *F* = .399, *p* = .672; LLAMA_E, *F* = 1.160, *p* = .318; LLAMA_F, *F* = 1.567, *p* = .215).

As a further step towards answering RQ5, a Pearson product moment correlation was calculated between implicit knowledge, explicit knowledge, language aptitude, and working memory. The results are shown in Table 4.31.

Table 4.31 Correlations (Pearson *r*): Language aptitude, working memory, implicit knowledge, and explicit knowledge

	BSD test	**Implicit knowledge**	**Explicit knowledge**
LLAMA test	.33** *p* = .00	.13 *p* = .23	.20 *p* = .06
BSD test		.17 *p* = .10	.07 *p* = .51
Implicit knowledge			.22* *p* = .04

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Table 4.31 shows that working memory correlated significantly with language aptitude (r = .33), while non-significant correlations between the ID variables and the scores of implicit and explicit knowledge were obtained for the sample as a whole, though the association between language aptitude and explicit knowledge approached significance (p = .057). As discussed in section 2.3, different levels of proficiency may show different patterns of correlations with language aptitude (Yalçin & Spada, 2016; Hummel, 2007) and working memory (Harrington & Sawyer, 1992; Kormos & Sáfár, 2008; Linck & Weiss, 2011). Taking this into consideration, individual correlations were run for each group of participants between ID variables (i.e. language aptitude and working memory), and their implicit and explicit knowledge scores as displayed in Table 4.32.

Table 4.32 Correlations (Pearson *r*): Language aptitude, working memory, implicit knowledge, and explicit knowledge by level groups

Level 5	**BDS test**	**Implicit knowledge**	**Explicit knowledge**
LLAMA test	.45** p = .01	.31 p = .10	.18 p = .33
BDS test		.37* p = .04	.19 p = .33
Implicit knowledge			.14 p = .47
Level 7	**BDS test**	**Implicit knowledge**	**Explicit knowledge**
LLAMA test	.04 p = .85	.44* p = .02	.06 p = .75
BDS test		.21 p = .27	-.07 p = .73
Implicit knowledge			.22 p = .25
Level 9	**BDS test**	**Implicit knowledge**	**Explicit knowledge**
LLAMA test	.43** p = .02	-.32 p = .09	.23 p = .23
BDS test		-.01 p = .95	-.03 p = .89

Implicit knowledge	.12 p = .54

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Table 4.32 indicates that language aptitude correlates significantly with working memory in Level 5 and Level 9, but not in Level 7. With regard to the association between ID variables and implicit and explicit knowledge, two significant moderate correlations were found between working memory and implicit knowledge in Level 5, and language aptitude and implicit knowledge in Level 7. It is interesting to notice that no significant correlations were found between the two ID factors and explicit knowledge.

Additional correlations were run between implicit knowledge, explicit knowledge, and the sub-components of the LLAMA test for the whole cohort of participants and for each individual group.

Table 4.33 Correlations (Pearson *r*): LLAMA test sub-components, implicit knowledge, and explicit knowledge for the whole cohort of participants

	Implicit knowledge	**Explicit knowledge**
LLAMA_B (Vocabulary learning)	.10 p = .31	.26** p = .01
LLAMA_D (Sound recognition)	.02 p = .83	.12 p = .27
LLAMA_E (Sound-symbol correspondence)	.09 p = .39	.08 p = .46
LLAMA_F (Grammatical inferencing)	.12 p = .25	.15 p = .17

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Table 4.33 shows that LLAMA_B (Vocabulary learning) correlates significantly with explicit knowledge (r = .26), while non-significant correlations between the other sub-components of the

LLAMA tests and explicit and implicit knowledge were obtained for the sample as a whole. As discussed in the Literature Review chapter, different sub-components of the LLAMA tests may show different patterns of correlations with explicit and implicit knowledge of a number of grammar points (Forsberg & Sandgreen, 2013; Yilmaz, 2012; Yalçin & Spada, 2016; see Li's (2015) meta-analysis for an overview).

4.8 RQ6a. Do language learning aptitude and working memory predict learners' explicit and implicit knowledge of difficult grammar points?

In light of some research findings on the role the ID variables language learning aptitude (Robinson, 1997; Yalçin & Spada, 2016) and working memory (Gilabert & Muñoz, 2010; Serafini & Sanz, 2015) may play in predicting learners' explicit and implicit knowledge of difficult grammar points, a specific hypothesis was formulated.

Hypothesis 1: Language learning aptitude and/or working memory will predict explicit and implicit knowledge of difficult grammar points, but not easy grammar points.

In order to test Hypothesis 1, two linear multiple regression analyses were conducted to see whether language learning aptitude and working memory predict learners' performance on the explicit and implicit measures of difficult grammar points; one linear multiple regression analysis was conducted for each dependent variable: MLK difficult grammar points (explicit knowledge) and EI/ON combined scores for difficult grammar points (implicit knowledge). In these analyses, L2 proficiency was included to control for any impact it might have on learners' performance on implicit and explicit measures besides their language aptitude and working memory. Pearson product moment correlations between variables are shown in Table 4.34.

Table 4.34 Correlations (Pearson *r*): L2 proficiency, language aptitude, working memory and explicit knowledge of difficult grammar points

	LLAMA test	**BDS test**	**MLK_difficult**
L2 proficiency	.09 *p* = .39	.05 *p* = .63	.26** *p* = .01
LLAMA test		.33** *p* = .00	.19 *p* = .07
BDS test			.02 *p* = .85

Note. MLK_difficult = explicit knowledge of difficult grammar points
* Correlation is significant at the 0.01 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

Table 4.34 shows a significant weak correlation between L2 proficiency and explicit knowledge of difficult grammar points; a non-significant correlation was found between working memory and explicit knowledge of difficult grammar points and between language aptitude and explicit knowledge of difficult grammar points. This last correlation approached significance (r = .194, p = .067) and indicates that a number of students with high language aptitude have good explicit knowledge of the 7 difficult grammar points (indefinite article, 3rd person –*s* in the simple present tense, second conditional, yes/no questions, dative alternation, verb complements, relative clauses). With respect to the association between L2 proficiency and explicit knowledge of difficult grammar points, the result is as expected as demonstrated in other studies (Absi, 2014; Roehr, 2005), that is, the higher the L2 proficiency the better performance on difficult grammar points in terms of explicit knowledge.

To see whether language aptitude (as measured by the LLAMA test) and working memory (as measured by the BDS test) would predict explicit knowledge of difficult grammar points, but not easy grammar points as has been hypothesized above, a linear multiple regression analysis was run. The assumptions of (a) normal distribution of data (distribution of residuals), (b) homogeneity of variances, (c) linearity, and (d) multicollinearity were checked, and although there is some distinct curvature in some of the correlations as shown in the multiple scatterplot for predictor variables, all the other assumptions were met, and therefore, running the linear regression analysis was viable (Larson-Hall, 2010) (see Appendix K).

A sequential regression was conducted to examine the effects of the ID variables language aptitude and working memory. The variable L2 proficiency was included in the analysis as a control variable and was entered first followed by the variables language aptitude and working memory. Language aptitude was entered second because the LLAMA test, as a more integrated measure, can gauge the cognitive functions of language analysis, memory, and phonetic coding (Skehan, 1989). Working memory was entered last because this basic measure can further provide predictive power on explicit and implicit difficult grammar points. The order in which the explanatory variables were entered in the regression analysis was followed in all subsequent regression analyses. The results in Table 4.35 of the linear regression analysis include the R square, R square change, the standardized β coefficients, and the significance value.

Table 4.35 Results of sequential regression analysis between the IVs L2 proficiency, language aptitude, working memory and the DV explicit knowledge of difficult grammar points

Model	**Predictor variables**	**R Square**	**R Square Change**	**Standardized β coefficients**	**Sig.**
1	L2 proficiency	.070	.070	.264	.01**
2	Language aptitude	.099	.029	.172	.10
3	Working memory	.102	.003	-.056	.61

Note. IVs = independent variables, DV = dependent variable
**Correlation is significant at the 0.01 level (2-tailed)

Table 4.35 shows that the variable L2 proficiency added the most explanatory power to the model, accounting for 7 percent of the variance of explicit knowledge of difficult grammar points (R^2 = .07). The variable language aptitude accounted for 3 percent of the variance, but it was statistically non-significant, and working memory accounted for 0.3 percent of the variance and it was statistically non-significant as well. Model 3, with all three predictors, accounted for

10 percent of the variance in explicit knowledge of difficult grammar points, but the only statistical predictor was L2 proficiency. Thus, the explanatory variables language aptitude and working memory do not significantly predict explicit knowledge of the targeted difficult grammar points for the cohort of participants as a whole, disconfirming hypothesis 1.

The same overall analysis was carried out for the easy grammar points as an additional check of hypothesis 1, for which no significant prediction was expected. Pearson product moment correlations between variables are shown in Table 4.36.

Table 4.36 Correlations (Pearson *r*): L2 proficiency, language aptitude, working memory and explicit knowledge of easy grammar points

	LLAMA test	**BDS test**	**MLK_easy**
L2 proficiency	.09 *p* = .20	.05 *p* = .32	.00 *p* = .49
LLAMA test		.33** *p* = .00	.18* *p* = .04
BDS test			.12 *p* = .13

Note. MLK_easy = explicit knowledge of easy grammar points
** Correlation is significant at the 0.01 level (2-tailed)
*Correlation is significant at the 0.05 level (2-tailed)

Table 4.36 shows a significant weak association between language aptitude and explicit knowledge of easy grammar points. No significant correlation was found between the scores of the working memory test and explicit knowledge.

To see whether language aptitude (as measured by the LLAMA test) and working memory (as measured by the BDS test) would predict explicit knowledge of easy grammar points, a linear multiple regression analysis was carried out. The assumptions of (a) normal distribution of data (distribution of residuals), (b) homogeneity of variance, (c) linearity, and (d) multicollinearity were checked, and although there is some distinct curvature in some of the correlations as shown in the multiple scatterplot for predictor variables, all the other assumptions were met, and therefore, running the linear regression analysis was viable (Larson-Hall, 2010) (see Appendix L). The results are displayed in Table 4.37.

Table 4.37 Results of sequential regression analysis between the IVs L2 proficiency, language aptitude, working memory and the DV explicit knowledge of easy grammar points

Model	Predictor variables	R Square	R Square Change	Standardized β coefficients	Sig.
1	L2 proficiency	.000	.000	.003	.98
2	Language aptitude	.032	.032	.180	.09
3	Working memory	.036	.004	.066	.56

Note. IV = independent variable, DV = dependent variable

Table 4.37 shows that the variable language aptitude was the strongest individual predictor, accounting for 3 percent of the variance (R^2 = .03) in learners' level of explicit knowledge of easy grammar points, followed by working memory, which accounts for a further 0.4 percent. The model including the three predictors, accounted for 4 percent of the variance in explicit knowledge of easy grammar points but none of the explanatory variables were statistically significant. Thus, the explanatory variables language aptitude and working memory do not significantly predict explicit knowledge of easy grammar points for the cohort of participants as a whole.

Another sequential linear regression analysis was conducted (assumptions were checked and met in full — see Appendix M) to find out whether language aptitude and working memory would predict implicit knowledge of difficult grammar points. Pearson product moment correlations between variables are shown in Table 4.38.

Table 4.38 Correlations (Pearson *r*): L2 proficiency, language aptitude, working memory and implicit knowledge of difficult grammar points

	LLAMA test	**BDS test**	**EI/ON_difficult**
L2 proficiency	.09 *p* = .39	.05 *p* = .63	.58** *p* = .00
LLAMA test		.33** *p* = .00	.19 *p* = .07
BDS test			.17 *p* = .11

Note. EI/ON_difficult = implicit knowledge of difficult grammar points
*Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

Table 4.38 shows an association of medium strength between L2 proficiency and implicit knowledge for the targeted difficult grammar points. This was probably due in part to L2 proficiency being measured using learners' narratives of the short story. Like the results for explicit knowledge of difficult grammar points, no significant associations were found between working memory and implicit knowledge of difficult grammar points or between language aptitude and implicit knowledge of difficult grammar points. The latter association approached significance (r = .190, p = .073), however, indicating that a number of students with high language aptitude have good implicit knowledge of the 7 difficult grammar points (indefinite article, 3rd person –*s* in the simple present tense, second conditional, yes/no questions, dative alternation, verb complements, relative clauses).

The results for the sequential regression analysis for implicit knowledge of difficult grammar points are displayed in Table 4.39.

Table 4.39 Results of sequential regression analysis between the IVs L2 proficiency, language aptitude, working memory and the DV implicit knowledge of difficult grammar points

Model	Predictor variables	R Square	R Square Change	Standardized β coefficients	Sig.
1	L2 proficiency	.338	.338	.582	.00**
2	Language aptitude	.357	.019	.138	.11
3	Working memory	.368	.010	.109	.24

**Correlation is significant at the 0.01 level (2-tailed)

As in the sequential regression analysis for the explicit difficult grammar points, Table 4.40 depicts that the variable L2 proficiency added the most explanatory power to the model, accounting for 34 percent of the variance ($R^2 = .34$) in learners' implicit knowledge of difficult grammar points. The predictor language aptitude accounted for 1 percent of the variance, but it was statistically non-significant, and working memory accounted for 2 percent of the variance and it was statistically non-significant as well. Model 3, with all three predictors, accounted for 37 percent of the variance in implicit knowledge of difficult grammar points, but the only statistical predictor was L2 proficiency. Thus, the explanatory variables language aptitude and working memory do not significantly predict implicit knowledge of the targeted difficult grammar points for the cohort of participants as a whole, disconfirming hypothesis 1.

With respect to implicit knowledge of easy grammar points, a sequential regression analysis was also planned (assumptions were checked and met—see Appendix N) as an additional check of hypothesis 1. Pearson product moment correlations between variables are shown in Table 4.40.

Table 4.40 Correlations (Pearson *r*): L2 proficiency, language aptitude, working memory and implicit knowledge of easy grammar points

	LLAMA test	BDS test	EI/ON_easy
L2 proficiency	.09 *p* = .00	.05 *p* = .32	.51** *p* = .00
LLAMA test		.33** *p* = .00	.03 *p* = .39
BDS test			.14 *p* = .10

Note. EI/ON_easy = implicit knowledge for easy grammar points
** Correlation is significant at the 0.01 level (2-tailed)

Table 4.40 shows a significant moderate association between L2 proficiency and implicit knowledge for the targeted easy grammar points, and as argued in the association between L2 proficiency and implicit knowledge of difficult grammar points, it is likely that such a correlation was due to the type of evaluation employed in assessing the L2 proficiency of learners. With respect to the other associations, no significant associations were found between language aptitude and implicit knowledge of easy grammar points, nor between working memory and implicit knowledge of easy grammar points. Despite these findings, a regression analysis was conducted to find out to what extent L2 proficiency would predict implicit knowledge of easy grammar points.

The results for the sequential regression analysis for implicit knowledge of easy grammar points are displayed in Table 4.41.

Table 4.41 Results of sequential regression analysis between the IVs L2 proficiency, language aptitude, working memory and the DV implicit knowledge of easy grammar points

Model	Predictor variables	R Square	R Square Change	Standardized β coefficients	Sig.
1	L2 proficiency	.261	.261	.511	.00**
2	Language aptitude	.261	.000	-.018	.85
3	Working memory	.276	.015	.130	.19

**Correlation is significant at the 0.01 level (2-tailed)

Table 4.41 shows that the variable L2 proficiency added the most explanatory power to the model ($R^2 = .26$) when it was added before language aptitude and working memory. After all the other variables were added, language aptitude accounted for zero percentage of the variance, and working memory accounted for 1.5% of the variance and it was statistically non-significant as well. Model 3, with all three predictors, accounted for 28% of the variance in implicit knowledge of easy grammar points, but the only statistical predictor was L2 proficiency. The explanatory variables language aptitude and working memory do not significantly predict implicit knowledge of easy grammar points.

As it was argued in Section 4.7, different sub-components of the LLAMA tests may show different patterns of correlations with explicit and implicit knowledge of a number of grammar points (Forsberg & Sandgreen, 2013; Yilmaz, 2012; Yalçin & Spada, 2016). Taking this into consideration, four more linear multiple regression analyses were conducted to see whether different aptitude components predict learners' performance on the explicit and implicit measures of difficult and easy grammar points; one linear multiple regression analysis was conducted for each dependent variable: MLK difficult grammar points, MLK easy grammar points, EI/ON combined scores for difficult grammar points, and EI/ON combined scores for easy grammar points. In these analyses, as in the previous multiple regressions, L2 proficiency was included to control for any impact it might have on learners' performance on implicit and explicit scores for difficult and easy grammar points besides their language aptitude abilities (aptitude components). Pearson product moment correlations between variables are shown in Table 4.42.

Table 4.42 Correlations (Pearson *r*): LLAMA test sub-components and explicit knowledge of difficult and easy grammar points

	MLK_difficult	MLK_easy
L2 proficiency	.26** p = .01	.00 p = .49
LLAMA_B (Vocabulary learning)	.24** p = .01	.23* p = .02
LLAMA_D (Sound recognition)	.07 p = .26	.16 p = .06
LLAMA_E (Sound-symbol correspondence)	.13 p = .11	.02 p = .44
LLAMA_F (Grammatical inferencing)	.11 p = .16	.17 p = .06

** Correlation is significant at the 0.01 level (2-tailed)
* Correlation is significant at the 0.05 level (2-tailed)

Table 4.42 shows that L2 proficiency and LLAMA_B (Vocabulary learning) correlated significantly with explicit knowledge of difficult grammar points (r = .26 and r = .24 respectively), while non-significant correlations between the other sub-components of the LLAMA tests and explicit knowledge of difficult grammar points were obtained for the sample as a whole. Similarly, LLAMA_B (Vocabulary learning) correlated significantly with explicit knowledge of easy grammar points, and LLAMA_D (Sound recognition) and LLAMA_F (Grammatical inferencing) approached significance.

To find out whether the sub-components of the LLAMA test would predict explicit knowledge of difficult and/or easy grammar points multiple regression analyses were carried out. The assumptions of (a) normal distribution of data (distribution of residuals), (b) homogeneity of variances, (c) linearity, and (d) multicollinearity were met. The results are displayed in Tables 4.43 and 4.44 for the sequential regression analyses between the sub-components of the LLAMA tests and explicit knowledge of difficult and easy grammar points.

Table 4.43 Results of sequential regression analysis between the IVs L2 proficiency, sub-components of LLAMA tests and the DV explicit knowledge of difficult grammar points

Model	Predictor variables	R Square	R Square Change	Standardized *β* coefficients	Sig.
1	L2 proficiency	.070	.070	.264	.01
2	LLAMA_B (Vocabulary learning)	.112	.042	.208	.04
3	LLAMA_D (Sound recognition)	.112	.000	-.007	.94
4	LLAMA_E (Sound-symbol correspondence)	.116	.004	.067	.54
5	LLAMA_F (Grammatical inferencing)	.116	.000	-.023	.85

Note. IV = independent variable, DV = dependent variable

Table 4.43 shows that the variable L2 proficiency was the strongest individual predictor, as expected, accounting for 7 percent of the variance (R^2 = .07) in learners' level of explicit knowledge of difficult grammar points, followed by vocabulary learning, which accounts for a further 4 percent. The model including the five predictors accounted for 12 percent of the variance in explicit knowledge of difficult grammar points but only the explanatory variables L2 proficiency and vocabulary learning were statistically significant.

As for the sequential regression analysis between the sub-components of the LLAMA tests and explicit knowledge of easy grammar points, the results are displayed in Table 4.44.

Table 4.44 Results of sequential regression analysis between the IVs L2 proficiency, sub-components of LLAMA tests and the DV explicit knowledge of easy grammar points

Model	Predictor variables	R Square	R Square Change	Standardized *β* coefficients	Sig.
1	L2 proficiency	.000	.000	.003	.97
2	LLAMA_B (Vocabulary learning)	.053	.053	.233	.03
3	LLAMA_D (Sound recognition)	.065	.012	.112	.30
4	LLAMA_E (Sound-symbol Correspondence)	.072	.007	-.090	.42
5	LLAMA_F (Grammatical inferencing)	.081	.009	.116	.36

Note. IV = independent variable, DV = dependent variable

In this regression analysis, the variable L2 proficiency did not predict learners' level of explicit knowledge of easy grammar points; vocabulary learning was the strongest individual predictor, accounting for 5 percent of the variance (R^2 = .05) in learners' level of explicit knowledge of easy grammar points. The model including the five predictors accounted for 8 percent of the variance in explicit knowledge of easy grammar points but only the explanatory variable vocabulary learning was statistically significant.

Two more regression analyses were conducted to see whether the sub-components of the LLAMA tests would predict implicit knowledge of difficult and/or easy grammar points. The assumptions of (a) normal distribution of data (distribution of residuals), (b) homogeneity of variances, (c) linearity, and (d) multicollinearity were met. The results are displayed in Tables 4.45 and 4.46 for the sequential regression analyses between the sub-components of the LLAMA tests and implicit knowledge of difficult and easy grammar points.

Table 4.45 Results of sequential regression analysis between the IVs L2 proficiency, sub-components of LLAMA tests and the DV implicit knowledge of difficult grammar points

Model	Predictor variables	R Square	R Square Change	Standardized β coefficients	Sig.
1	L2 proficiency	.338	.338	.582	.00
2	LLAMA_B (Vocabulary learning)	.340	.002	.047	.59
3	LLAMA_D (Sound recognition)	.340	.000	-.006	.95
4	LLAMA_E (Sound-symbol correspondence)	.361	.021	.155	.10
5	LLAMA_F (Grammatical inferencing)	.369	.008	.107	.32

Note. IV = independent variable, DV = dependent variable

Table 4.45 shows that the variable L2 proficiency was the strongest individual predictor, accounting for 34 percent of the variance (R^2 = .34) in learners' level of implicit knowledge of difficult grammar points. The model including the five predictors accounted for 37 percent of the variance in implicit knowledge of difficult grammar points but only the explanatory variable L2 proficiency was statistically significant.

With respect to the sequential regression analysis between the sub-components of the LLAMA tests and implicit knowledge of easy grammar points, the results are displayed in Table 4.46.

Table 4.46 Results of sequential regression analysis between the IVs L2 proficiency, sub-components of LLAMA tests and the DV implicit knowledge of easy grammar points

Model	Predictor variables	R Square	R Square Change	Standardized β coefficients	Sig.
1	L2 proficiency	.261	.261	.511	.00
2	LLAMA_B (Vocabulary learning)	.261	.000	-.016	.86
3	LLAMA_D (Sound recognition)	.267	.006	-.079	.40
4	LLAMA_E (Sound-symbol correspondence)	.267	.000	.012	.90
5	LLAMA_F (Grammatical inferencing)	.268	.001	.047	.68

Note. IV = independent variable, DV = dependent variable

Like the regression analysis for implicit knowledge of difficult grammar points, Table 4.46 shows that the variable L2 proficiency was the strongest individual predictor, accounting for 26 percent of the variance ($R^2 = .26$) in learners' level of implicit knowledge of easy grammar points. The model including the five predictors accounted for 27 percent of the variance in implicit knowledge of easy grammar points but only the explanatory variable L2 proficiency was statistically significant.

In sum, when regression analyses were run for the sub-components of the LLAMA test, LLAMA_B (Vocabulary learning) was a significant predictor for explicit knowledge of both difficult and easy grammar points; none of the sub-components of language aptitude significantly predicted implicit knowledge of difficult or easy grammar points. On the other hand, the overall findings in this section show that the ID variables language aptitude and working memory do not significantly predict explicit and implicit knowledge of difficult or easy grammar points. Similar findings were obtained in Serafini and Sanz' (2015) study for their whole sample of participants (see section 2.4.4). In order to find out what role language aptitude and working memory would play at each individual proficiency level in the present study (intermediate, upper-intermediate, advanced) (Kormos & Sáfár, 2008; Li, 2014; Serafini & Sanz, 2015) , an analysis by level group was conducted.

4.9 RQ6b. Do language learning aptitude and working memory predict learners' explicit and implicit knowledge of difficult grammar points in participants at different levels?

There have been differing positions on the contexts in which language aptitude and working memory may play a role in SLA at different levels of L2 proficiency. It has been suggested that language aptitude may only be important at lower levels of language proficiency (Hummel, 2009; Li, 2014; Skehan, 2012); this same suggestion has been made for the variable working memory (Kormos & Sáfár, 2008; Linck & Weiss, 2011; Serafini & Sanz, 2015).

Hypothesis 2: Language learning aptitude and/or working memory will predict the explicit and implicit knowledge of difficult and easy grammar points in the Level 5 group, but not the Level 9 group, with the Level 7 group falling in between.

Further linear regressions were conducted to see whether language learning aptitude and working memory can predict learners' performance on the explicit and implicit measures of difficult or easy grammar points for each group of participants (levels 5, 7, and 9). Pearson product moment correlations were run (the assumption of normal distribution was met, see Table 3.24) between the ID variables (language aptitude and working memory) and explicit knowledge of difficult grammar points for the three level groups; no significant correlations were found for any of the groups as shown in Table 4.48. Based on these findings, no regression analysis was conducted.

Table 4.48 Correlations (Pearson *r*): L2 proficiency, language aptitude, working memory and explicit knowledge of difficult grammar points by level group

Level 5	L2 proficiency	LLAMA test	BDS test
MLK_difficult	.15 *p* = .22	.15 *p* = .22	.16 *p* = .20
Level 7			
MLK_difficult	.18 *p* = .18	.16 *p* = .21	-.17 *p* = .18
Level 9			
MLK_difficult	.13 *p* = .25	.18 *p* = .18	-.07 *p* = .35

Note. MLK_difficult = explicit knowledge of difficult grammar points
*Correlation is significant at 0.05 level (2-tailed)

Following the rationale for the statistical analyses carried out for the easy grammar points on the cohort of participants as a whole, a linear regression was planned to see whether language learning aptitude and working memory can predict learners' performance on the explicit measures of easy grammar points for each group of participants (levels 5, 7, and 9). Pearson product moment correlations were conducted between the ID variables and explicit knowledge of easy grammar points for the three level groups; no significant correlations for explicit knowledge of easy grammar points were found for any of the groups as shown in Table 4.49. Based on these findings, no regression analysis was conducted.

Table 4.49 Correlations (Pearson *r*): L2 proficiency, language aptitude, working memory and explicit knowledge of easy grammar points by level group

Level 5	L2 proficiency	LLAMA test	BDS test
MLK_easy	.07 *p* = .37	.21 *p* = .14	.20 *p* = .15
Level 7			
MLK_easy	-.09 *p* = .33	-.07 *p* = .36	.08 *p* = .33
Level 9			
MLK_easy	-.18 *p* = .17	.23 *p* = .11	.02 *p* = .45

Note. MLK_easy = explicit knowledge of easy grammar points
*Correlation is significant at 0.05 level (2-tailed)

Concerning implicit knowledge of difficult grammar points, a Pearson product moment correlation was run (the assumption of normal distribution was met) for the variables L2 proficiency, language aptitude, working memory and implicit knowledge of difficult grammar points for each group of participants (level 5, 7, and 9); Table 4.50 shows the correlations between the variables.

Table 4.50 Correlations (Pearson *r*): L2 proficiency, language aptitude, working memory and implicit knowledge of difficult grammar points by level group

Level 5	**L2 proficiency**	**LLAMA test**	**BDS test**
EI/ON_difficult	.29 *p* = .06	.36* *p* = .02	.25 *p* = .09
Level 7			
EI/ON _difficult	.72** *p* = .00	.46* *p* = .01	.20 *p* = .15
Level 9			
EI/ON_difficult	.55** *p* = .00	-.32* *p* = .045	.04 *p* = .43

Note. EI/ON_difficult = implicit knowledge of difficult grammar points
* Correlation is significant at 0.05 level (2-tailed)
** Correlation is significant at 0.01 level (2-tailed)

Table 4.50 displays significant correlations between L2 proficiency and implicit knowledge of difficult grammar points for Level 7 and Level 9, with the correlation for Level 5 approaching significance, and between language aptitude and implicit knowledge of difficult grammar points for the three Levels. Surprisingly, the correlation in Level 9 is negative, suggesting that the higher language aptitude on the learners' part, the lower the scores they obtained on the implicit measures of difficult grammar points.

A linear regression analysis was run (assumptions were checked and met—see Appendix O) to find out whether language aptitude (as measured by the LLAMA test) and working memory (as measured by the BDS test) would predict implicit knowledge of difficult grammar points in each group of participants. Table 4.51 shows the results of the regression analysis.

Table 4.51 Results of a sequential regression analysis between the IVs L2 proficiency, language aptitude, working memory and the DV implicit knowledge of difficult grammar points by level group

Model	Predictor variables	R Square	R Square Change	Standardized β coefficients	Sig.
Level 5					
1	L2 proficiency	.087	.087	.294	.12
2	Language aptitude	.202	.115	.340	.06
3	Working memory	.214	.012	.124	.53
Level 7					
1	L2 proficiency	.522	.522	.723	.00*
2	Language aptitude	.558	.036	.206	.15
3	Working memory	.568	.010	.100	.45
Level 9					
1	L2 proficiency	.301	.301	.549	.01**
2	Language aptitude	.321	.020	-.150	.38
3	Working memory	.345	.024	.172	.34

* Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

Table 4.51 shows that the variable L2 proficiency added the most explanatory power to the model in Level 7, accounting for 52 percent of the variance in learners' implicit knowledge of difficult grammar points and 30 percent in Level 9 (R^2 = .52 and .30 respectively), and it was statistically significant. Likewise, language aptitude added the most explanatory power to the model in Level 5, accounting for 12 percent of the variance (R^2 = .12), and it approached significance. Working memory does not significantly predict implicit knowledge of the targeted difficult grammar points for any group of participants.

With regard to implicit knowledge of easy grammar points, a sequential regression analysis was run (assumptions were checked and met—see Appendix P). A Pearson product moment correlation was run for the variables L2 proficiency, language aptitude, working memory and implicit knowledge of easy grammar points for each group of participants (Level 5, 7, and 9); Table 4.52 shows the correlations between the variables.

Table 4.52 Correlations (Pearson *r*): L2 proficiency, language aptitude, working memory and implicit knowledge of easy grammar points by level group

Level 5	**L2 proficiency**	**LLAMA test**	**BDS test**
EI/ON_easy	.49* *p* = .01	.09 *p* = .32	.33* *p* = .04
Level 7			
EI/ON_easy	.49* *p* = .01	.36* *p* = .03	.20 *p* = .15
Level 9			
EI/ON_easy	.60** *p* = .00	-.27 *p* = .07	-.05 *p* = .39

* Correlation is significant at 0.05 level (2-tailed)
** Correlation is significant at 0.01 level (2-tailed)

Table 4.52 displays significant correlations between L2 proficiency and implicit knowledge of easy grammar points for the three Levels, and between working memory and implicit knowledge for Level 5 and language aptitude and implicit knowledge for Level 7. The latter two associations are at a moderate level of strength. The counter-intuitive negative coefficient between language aptitude and implicit knowledge of easy grammar points in Level 9 approached significance, which shows the same pattern as the negative correlation between language aptitude and implicit knowledge of difficult grammar points in Level 9 in Table 4.51.

A linear regression analysis was run (assumptions were checked and met—see Appendix P) to find out whether language aptitude (as measured by the LLAMA test) and working memory (as measured by the BDS test) would predict implicit knowledge of easy grammar points in each group of participants. Table 4.53 shows the results of the regression analysis.

Table 4.53 Results of a sequential regression analysis between the IVs L2 proficiency, language aptitude, working memory and the DV implicit knowledge of easy grammar points by level group

Model	Predictor variables	R Square	R Square Change	Standardized *β* coefficients	Sig.
Level 5					
1	L2 proficiency	.239	.239	.488	.01*
2	Language aptitude	.241	.003	.053	.75
3	Working memory	.364	.123	.393	.03*
Level 7					
1	L2 proficiency	.237	.237	.487	.01*
2	Language aptitude	.271	.034	.200	.27
3	Working memory	.288	.018	.134	.43
Level 9					
1	L2 proficiency	.354	.354	.595	.00**
2	Language aptitude	.361	.006	-.085	.81
3	Working memory	.362	.001	.038	.63

Note. Dependent variable: implicit knowledge of easy grammar points
* Correlation is significant at the 0.05 level (2-tailed)
** Correlation is significant at the 0.01 level (2-tailed)

Table 4.53 shows that the variable L2 proficiency added the most explanatory power to the model in each of the three Levels, accounting for 24 percent ($R^2 = .24$) of the variance in learners' implicit knowledge of easy grammar points in Level 5, 24 percent ($R^2 = .24$) in Level 7 and 35 percent ($R^2 = .35$) in Level 9, and it was statistically significant. The ID working memory is the only variable that increased the explained variance from 24 to 36 percent in Level 5. That is, 12 percent ($R^2 = .12$) of the variability in learners' implicit knowledge of easy grammar points was predicted by working memory when L2 proficiency was controlled. The explanatory variable language aptitude does not significantly predict implicit knowledge of the targeted easy grammar points for any group of student participants.

4.10 RQ7. Is L2 use outside the classroom related to participants' performance on the measures of explicit and implicit knowledge of the targeted grammar points?

Further statistical analyses were run for the background variables use of English at home, use of English at work, and attendance at the university self-access centre to see to what extent these three variables correlate with the participants' performance on the implicit and explicit measures. The use of L2 outside the classroom appears to help in the L2 development, especially in instructed contexts where poor environments for language use prevail (Bialystok, 1979; Sorace, 1985). Put differently, learners using the L2 in real contexts for spontaneous communication may activate implicit learning mechanisms. These variables were measured in times per week. Recall that students must do activities at a self-access centre for a total of 8 hours representing 10 percent of the final grade on their English course.

The assumption of normal distribution of data was not met for any background variable. Therefore, a Spearman rank order correlation was conducted between the participants' background variables and their implicit knowledge (EI/ON combined scores) and explicit knowledge scores (MLK scores). One significant correlation was found as shown in Table 4.58.

Table 4.54 Correlations between the use of English at home and work, attendance to CAADI and implicit and explicit scores.

Variable	N	Implicit knowledge	Explicit knowledge
Use of English at home	13	.16 *p* = .13	-.07 *p* = .50
Use of English at work	11	.16 *p* = .13	.27* *p* = .01
Attendance at self-access center	29	-.21 *p* = .05	-.16 *p* = .12

Note. Implicit knowledge = EI/ON combined scores
Explicit knowledge = MLK test scores
*. Correlations is significant at the 0.05 level (2-tailed)

Table 4.58 shows a positive significant correlation between use of English at work and explicit knowledge of targeted grammar points, indicating that the more English the learners use at work, the higher the score on the MLK test is, or vice versa, though the strength of the correlation is weak. Finding a correlation between these two variables is counter-intuitive, that is, a correlation was expected between use of English at work and implicit knowledge because the use of English at work usually involves more aural/oral skills; however, the result appears to indicate that English at work may refer to learners working as teachers, as indicated in Section 3.2.1. On the other hand, the trend towards a negative association between attendance at the self-access centre and implicit knowledge of targeted grammar points is unexpected considering that the more hours learners spend at the self-access centre either doing reading, writing, or listening activities, or doing grammar exercises should help them develop their knowledge of the grammar points in question. A feasible interpretation is that the negative association indicates that the weaker learners were the ones who attended the self-access centre. With respect to the non-significant correlation between use of English at home and implicit and explicit knowledge, 13 participants reported having used English at home. From these cases, eight cases reported having used English at home 1, 2, or 3 times per week, one 4 times, and four 7 times. The number of cases per times per week shows some discrimination of the use of English at home, but the sample size (N = 13) may be too small to show a significant result.

CHAPTER FIVE: DISCUSSION

In the present study 13 targeted grammar points were judged by learners (*N* = 90) and teachers (*N* = 26) as easy grammar points (plural of nouns, simple past tense (-*ed* form), modal verbs, *many* vs. *much*, comparative adjectives, *since* vs. *for*) and difficult grammar points (indefinite article, simple present tense (3rd person –*s*), verb complements, second conditional (*if*-clauses), *yes/no* questions, dative alternation, relative clauses). The student participants were tested on these grammar points by means of an explicit measure (MLK test) and two implicit measures (EI test and ON test). It was hypothesized that the ID variables language learning aptitude (as measured by the LLAMA test) and/or working memory (as measured by the BDS test) would predict learners' explicit and implicit knowledge of difficult grammar points, but not easy grammar points. Following initial analyses, a second hypothesis was formulated in the sense that these two ID variables would predict explicit and implicit knowledge of difficult and easy grammar points in the lower level group (Level 5) but not the higher levels (Level 7 and 9). Furthermore, L2 proficiency was included in the statistical analyses as a control variable.

5.1 Teachers' and learners' perceived learning difficulty of targeted grammar points

The categorization of the 13 selected grammar points in the present study was approached by asking the learner participants (*N* = 90) and teacher participants (*N* = 26) to complete a difficulty judgement questionnaire including a 5-point scale (very easy – easy – moderate – difficult – very difficult) (DeKeyser, 2003) based on their experience on teaching and learning the language (teachers) and according to their experience in learning the language (learners). The learner participants judged eight grammar points as easy and five as difficult, while the teacher participants classified six grammar points as easy and seven as difficult. When comparing both types of judgments, both learners and teachers perceived the same grammar points as easy or difficult, with the exception of two grammar points: 3rd person –*s* and yes/no questions are grouped as easy by the learners and as difficult by the teachers. These two grammar points were assigned to the difficult category by the researcher for three reasons: (1) a Spearman's rank order correlation was run between the learners' and the teachers' mean difficulty scores to find out whether there was any correlation. There was a significant and strong positive correlation between the two sets of mean scores (rho = .742, $p < .01$), (2) one study used the expert judgement of experienced L2 teachers to categorize a number of grammar points as easy or difficult (Robinson, 1996) and, (3) in a study that was conducted prior to this study (Rodríguez Silva & Roehr-Brackin, 2016) the judgment of 11 teacher participants showed a tendency towards successful prediction of learners' performance on both explicit and implicit L2 measures.

The concept of learning difficulty of grammar points can be approached from a subjective perspective (processing difficulty or cognitive difficulty) and/or an objective perspective (structural complexity of the linguistic construction) (DeKeyser, 2003; Housen and Simoens, 2016). Studies focusing on one approach or the other to categorize a number of grammar points into easy or difficult may make no distinction between implicit and explicit knowledge of grammar points, or they may disregard factors (R. Ellis, 2008; Dörnyei, 2005; Graus and Coppen, 2015) such as perceptual salience, communicative redundancy, conceptual complexity, and technicality of metalanguage (just to mention a few) that pertain to implicit and/or explicit learning difficulty.

In the current study, the fact that learners did not categorize any of the grammar points as difficult or very difficult, and that none of the grammar points were classified as very easy or very difficult by teachers may be an indication that other factors are at play. Similar patterns of results have been found in other studies (Absi, 2014; Huang, 2012; Rodríguez Silva & Roehr-Brackin, 2016; Thepeseenu & Roehr, 2013). One factor that may explain why learners in the present study judged most grammar points as easy might have to do with the way they are instructed, that is, it is possible that they may have spent more time in trying to understand the form of grammatical structures (Shiu, 2011) than their meaning. It may be also possible that learners did not have the opportunity to use the grammar points in social contexts, and not just in the linguistic context of the classroom. In other words, it is likely that learners used the grammar points inside the classroom only. This means that the prolonged exposure to the explicit instruction of the grammar points may be one factor explaining why learners judged the grammar points as less difficult than did the teachers, but whether learners actually find the selected grammar points easy to use is a different issue, as discussed in section 5.2.

A possible explanation for the teachers' categorization of the grammar points may be that they were aware that the L2 learning difficulty of the targeted grammar points is not very easy or easy as learners claimed for most grammar points; in fact, teachers classified most grammar points as moderate and difficult. These results are commensurate with the categorization of grammar points in other studies (Absi, 2014; Huang, 2012). Put differently, teachers in the present study may not have the theoretical knowledge of the characteristics of the 13 linguistic structures as implicit knowledge (e.g. frequency, perceptual salience, communicative redundancy), the characteristics of the linguistic descriptions as explicit knowledge (e.g. schematicity, conceptual complexity, technicality of metalanguage) (DeKeyser, 2005; R. Ellis, 2006), or the characteristics of learners (IDs), but their intuitions on how learners learn and use an L2 (Scheffler, 2011) show evidence of their L2 learning and teaching experience. In addition to this, teachers are witness of the learning difficulty and achievement of their learners, and this gives them a broader perspective of how difficult each grammar point may be for their learners.

With regard to the finding of the strong, positive correlation between the difficulty judgements made by the teachers and learners, this is as expected considering that they might have used similar judgement criteria given that both share the same teaching and learning setting.

Nevertheless, this interpretation is only speculative because no interviews or classroom observations were made. A similar result was found in Rodríguez Silva and Roehr-Brackin (2016) despite the smaller number of student participants (*N* = 30) and teacher participants (*N* = 11). This is explained in the sense that the study was conducted in the same educational setting of the current study.

5.2 Learners' performance on the implicit and explicit measures

In the present study, the results reveal that the overall explicit knowledge score (i.e. MLK test score; mean % = 67) was similar to the implicit knowledge score (i.e. EI/ON combined score; mean % = 66).

It was found that learners performed significantly better on easy grammar points (mean % = 75) than difficult grammar points (mean % = 61). Significant differences with large effect sizes in each group of participants with respect to easy vs. difficult explicit grammar points were found. With regard to the explicit knowledge scores by level groups, learners in Level 7 (mean % = 70) and 9 (mean % = 71) performed significantly better than learners in Level 5 (mean % = 61).

A similar pattern of results was found for implicit knowledge scores. Learners performed significantly better on easy grammar points (mean % = 70) than difficult grammar points (mean % = 62). Like the explicit knowledge scores, significant differences were found between the scores for easy and difficult grammar points in each level group. Regarding the implicit knowledge scores by level groups, learners in Level 7 (mean % = 68) and 9 (mean % = 69) showed a trend towards outperforming learners in Level 5 (mean % = 61).

A Pearson product moment correlation for the overall scores between the measures of implicit and explicit knowledge showed a significant though weak positive correlation ($r = .22$), while a Spearman's rank order correlation between implicit and explicit knowledge scores by grammar point showed a negative correlation approaching significance (rho = -.54, *p* = .058). It is worth discussing that implicit and explicit knowledge diverge significantly in Level 9, but not in the other Levels. In other words, it is the highest-level group that appears to drive the marginal negative correlation between implicit and explicit scores by grammar point.

Turning to the implicit and explicit scores, learners' overall performance on the implicit (mean % = 66) and explicit (mean % = 67) measure seems to indicate that learners developed both implicit and explicit knowledge of the targeted grammar points to a similar extent; the positive correlation between implicit and explicit knowledge endorses this interpretation. This is further corroborated by the moderate positive correlation between EI and MLK test scores (see section 5.3). However, the negative correlation by grammar point between implicit and explicit knowledge appears to contradict such similarity in performance. On one hand, the significant correlation at a global level between implicit and explicit knowledge is in consonance with other studies (Absi, 2014, Akakura, 2014; R. Ellis, 2005, Rodríguez Silva & Roehr-Brackin, 2016) suggesting that instructed learners' implicit and explicit knowledge correlate if these types of knowledge are assessed by using measures testing a range of L2 constructions (Absi, 2014; Alipour, 2014; Scheffler & Cinciata, 2011). On the other hand, the trend towards a non-significant negative association between implicit and explicit knowledge by targeted grammar point indicates that learners appear to have developed implicit knowledge of certain grammar points and explicit knowledge of other grammar points, and vice versa, but not necessarily both implicit and explicit knowledge of the same grammar point. This finding is in concordance with R. Ellis' (2006) and Rodríguez Silva and Roehr-Brackin's (2016) findings, that is, the pattern of results suggests that either explicit or implicit knowledge was developed first and then the other type of knowledge for each targeted grammar point, particularly in learners with longer L2 experience, that is, learners in Level 9. In learners at lower levels (Level 5) this is less evident.

To illustrate that learners may not develop both types of knowledge simultaneously, in Krashen's terms "learned knowledge" (explicit knowledge) and "acquired knowledge" (implicit knowledge) (Krashen, 1981, 1982, 1985), the grammar point 3rd person *–s* in the simple present tense is an example (along with other grammar points) of such a situation in the present study. An analysis by level group (Level 5, 7, and 9) reveals that the participants found this grammatical construction relatively easy in terms of explicit knowledge (73%, 73%, and 83% respectively), and difficult in terms of implicit knowledge (41%, 49%, and 56% respectively). The low explicit vs. high implicit difficulty for learning and acquiring this grammar point has been discussed theoretically (Krashen 1982; Collins et al., 2009; DeKeyser, 2005; N. Ellis, 2006; Spada & Tomita, 2010) and researched empirically (Absi, 2014; Erlam, 2006; R. Ellis, 2005; Rodríguez Silva & Roehr-Brackin, 2016) in SLA in the past decades, and the theoretical and empirical results suggest that this particular grammar point is easier to learn explicitly than implicitly in an instructed setting, as evidenced in the present study. Similar patterns were found for the grammar points yes/no questions, since/for, many vs. much, modal verbs, and comparative adjectives. It is interesting to note that these grammar points, with the exception of yes/no questions, were categorized as easy suggesting that participants would find them easy to learn both explicit and implicitly.

Conversely, participants found the grammar point relative clauses relatively easy in terms of implicit knowledge (77%, 88%, and 85% respectively), and difficult in terms of explicit knowledge (41%, 55%, and 57% respectively), and similar patterns were found for the grammar points indefinite article and dative alternation. These grammar points, as opposed to the easy grammar points just discussed, were categorized as difficult suggesting that participants would find them difficult to learn both implicit and explicitly. This illustration of what grammar points learners found easy or difficult in terms of explicit or implicit knowledge suggests that learners categorised the grammar points based on their explicit knowledge as argued in section 5.1.

The question why some grammar points are easy in terms of explicit knowledge and difficult in terms of implicit knowledge and other grammar points are easy in terms of implicit knowledge and difficult in terms of explicit knowledge, as illustrated in the preceding paragraph, can be answered by considering a number of variables. Roehr and Gánem-Gutiérrez' (2009a) taxonomy (see Table 2.5 in section 2.2.2.3) presents nine variables that contribute to implicit and explicit learning difficulty of grammar points. Each individual variable (frequency, perceptual salience, communicative redundancy, opacity of form-meaning mapping, opacity of meaning-form mapping, schematicity, conceptual complexity, technicality of metalanguage, truth value) can help explain, to some extent, either the implicit or explicit learning difficulty of grammar points sharing similar (meta-)linguistic characteristics. A combination of two or more factors can further help explain such learning difficulty of grammar points as implicit or explicit knowledge. Thus, for instance, if a grammar point can be described by means of a metalinguistic rule which is relatively high in schematicity, relatively low in conceptual complexity, relatively high in truth value, and makes use of relatively non-technical metalanguage, explicit learning difficulty is low. To illustrate this, in the present study the grammar point 3rd person –*s* is high in schematicity because it is a general rule that applies to all verbs conjugated with the pronouns *he*, *she*, or *it*, its conceptual complexity is low because there is only one relation between two categories that need to be taken into consideration (i.e. 3rd singular person and verb in simple present tense), its technicality of metalanguage is low because the metalanguage (simple present tense) is non-technical and familiar to most learners, and its truth value is high because there are few exceptions to the rule.

Conversely, if a grammar point is described by a rule that is low in schematicity, relatively high in conceptual complexity, relatively low in truth value, and makes use of relatively technical metalanguage, explicit learning difficulty is high. To illustrate, in the current study the grammar point dative alternation is low in schematicity because the rule does not apply to all verbs, its conceptual complexity is high because there is more than one relation between categories that need to be taken into consideration, its technicality of metalanguage is high because the metalanguage is unfamiliar to most learners, and its truth value is low because there are a number of exceptions to the rule.

Similarly, if a grammar point has high frequency, high perceptual salience, low communicative redundancy, and low opacity, implicit learning difficulty is low. To illustrate, in the present study the grammar point indefinite article has high frequency because it is commonly heard in naturalistic input, its perceptual salience is medium because it is seldom stressed in spoken input, its communicative redundancy is low because the omission of the indefinite article affects the meaning of the utterance, and its opacity (one form, X meanings; one meaning, X forms) is low because if one wants to talk about a singular count noun for the first time an indefinite article is the form to use. Conversely, if a grammar point has low frequency, low perceptual salience, high communicative redundancy, and high opacity, implicit learning difficulty is high. To illustrate, in the current study the grammar point second conditional has low frequency because it is not commonly heard in naturalistic input, its perceptual salience is medium because the modal verb *would* in the main clause is usually contracted and seldom stressed, its communicative redundancy is high because an error in any of the clauses does not stop a speaker to get his or her message across, and its opacity (one form, X meanings; one meaning, X forms) is medium because the form of the verb in the dependent clause is also used in simple past tense.

It is also worth considering the role L1 plays in adult learners in the learning and acquisition of an L2. Put differently, “the very things that make a known language easy make a new language hard” (N. Ellis & Larsen-Freeman, 2006, p. 568). The interpretation of this may be that in the L1

(Spanish) of the learners in the present study, the present simple is an inflectional tense[7] which contrasts with the target language (English) with only one inflection in the simple present tense, and one possible explanation why learners in the present study omitted the third-person –*s* was because they possibly transferred entrenched constructions from their L1 corresponding to the new construction in L2. A second possible explanation can be that the final consonant clusters in verb conjugations in English do not exist in the learners' L1 (Spanish). A third possible explanation for such an omission, and closely related to the first explanation, is that L1 transfer and salience play an important role in L2 acquisition. From the notion of the definition of salience as "the general perceived strength of stimuli" (N. Ellis, 2006a, p. 16), it can be understood that the linguistic knowledge learners have of their L1 overshadowed the learning and use of L2 grammar points because either the third-person –*s* is not salient enough to be noticed in spontaneous speech, and/or prior learning of L1 constructions inhibited new learning of L2 linguistic constructions (N. Ellis, 2006b), particularly where structures differ between L1 and L2 (DeKeyser, 1998, 2005, 2016); N. Ellis (2006b) explains this situation as "the difficulties of adult L2 acquisition are a result of prior L1 learning, entrenchment, and transfer" (p. 185). In other words, learners in the current study possibly transferred entrenched constructions from their L1 corresponding to the new constructions in L2 such as the third-person –*s*, comparative adjectives, verb complements, and modal verbs.

In addition to this, it is likely that such a difference between these two languages in the sense of the conjugation of the verbs in simple present tense, as well as the one-syllable adjectives, the –*ing* form construction in verb complements, and the simple form of the main verb with modal verbs may place a heavy cognitive burden for L1 Spanish learners of L2 English to process and internalize these grammar points. Della Putta (2016, p. 220) refers to this phenomenon as

[7] Spanish (Mexican) grammar has five inflections for any verb conjugated with subject pronouns (see Schmitt, 2008).

"unlearning." In other words, "learners need to learn how to inhibit the automatic activation of an L1 grammar point in L2 contexts potentially triggering it" (Della Putta, 2016, p. 220). This represents a high cognitive demand for Spanish learners to use these grammar points communicatively, and this is not only reflected in the performance of learners in Level 5 but also in Level 7 and 9 in the present study.

If a comparison is made between R. Ellis' (2006) larger-scale study targeting 17 grammar points of L2 English and the present study, 11 out of the 17 grammar points are the same in both studies (relative clauses, dative alternation, *since* vs. *for*, modal verbs, comparative adjectives, *yes*/*no* questions, simple past tense (*-ed* ending), indefinite article, plural of nouns, 3rd person *–s* in the simple present tense, verb complements). Participants in both studies found the grammar points *yes*/*no* questions, comparative adjectives, 3rd person *–s*, and *since* vs. *for* explicitly easier, and dative alternation implicitly easier. No commonalities were found for the rest of the grammatical constructions, in fact, the participants in the current study found the rest of the grammar points (except modal verbs which showed a similar trend in R. Ellis' (2006) study) implicitly easier while in R. Ellis' (2006) study these grammar points showed the opposing trend. This may be due to the differences of the L1 of the learners (Spanish in the present study, and Chinese, Japanese, and Malaysian in R. Ellis' (2006) study).

Following DeKeyser's (1998, 2005, 2016) argument about the complexity of form-meaning mapping of some grammatical structures such as the 3rd person *–s*, comparative adjectives, and *yes/no* questions, the relationship between form and meaning may not be easy to process due to the non-salient suffix *–s* (3rd person *–s*), the bound morpheme *–er* (comparative adjectives), and the auxiliary verbs *do*/*does* (yes/no questions) which are non-existent in the L1 (Spanish) of the participants in the present study (see Dixon and Andújar, 1967, for comparative characteristics between English and Spanish). These differences between L1 and L2 suggest that what does not exist in learners' mother tongue may take learners more time to notice and consequently to learn implicitly, although explicit knowledge about these grammar points is developed without much difficulty (DeKeyser, 1998, 2005, 2016). Nevertheless, the commonalities found for the five grammar points further corroborate what was pointed out before that the development of both implicit and explicit knowledge for a number of grammar points may not happen concurrently.

5.3 Learners' performance on the EI test and the MLK test

Descriptive statistics and inferential statistics were calculated between the overall EI test score and MLK test score because the EI test includes all the grammar points of the current study. The results for these tests show that the overall MLK test score (i.e. mean % = 67) was higher than the EI test score (i.e. EI test score; mean % = 61).

Regarding the EI test scores, learners performed significantly better on easy grammar points (mean % = 64) than difficult grammar points (mean % = 58). Like the MLK test scores, significant differences were found between the scores for easy and difficult grammar points in each level group. Regarding the EI test scores by level groups, only learners in Level 9 (mean % = 66) outperformed learners in Level 5 (mean % = 56); no significant differences were found between Level 5 and Level 7. Furthermore, the overall scores for the targeted grammar points and the scores for the grammatical and ungrammatical items in this test are higher than the scores obtained in Erlam's (2006) study: her learners obtained 51 percent on the test as a whole, repeated 61 percent of grammatical items correctly, and corrected 39 percent of ungrammatical items. In contrast, the current study showed the results as follows: learners obtained 61 percent on the test as a whole, repeated 72 percent of grammatical items correctly, and corrected 49 percent of ungrammatical items. This reveals that learners in two different learning contexts (Mexican learners learning English in Mexico and Erlam's participants learning English in New Zealand, an English speaking country) scored differently on grammatical and ungrammatical grammar points. It is plausible that the Mexican learners performed higher at a global level and on both grammatical and ungrammatical grammar points than Erlam's learners due to their level of proficiency in English, that is, the current study did not include lower intermediate learners as Erlam's study did. Another possible explanation is that the prolongued exposure to the form of the grammatical structures may have activated not only the learners' explicit learning mechanisms but also their implicit learning mechanisms.

Learners' poorer performance on the EI test than the MLK test may be linked to the conjecture made in section 5.1 that learners may have spent more time in gaining explicit knowledge of the grammar points than learning the meaning of the structures and using the structures in spontaneous communication (implicit knowledge) in different contexts. Bearing this in mind, it can be argued that this type of instruction may have set their frames of mind to the study about the language (explicit knowledge) than the use of the language (implicit knowledge). On the other hand, that learners performed better on easy grammar points than difficult grammar points on the EI test endorses learners' and teachers' accurate judgements, and such performance can probably be interpreted as how close the L1-L2 distance is. In other words, the easy grammar points follow a similar singular-plural form (plural of nouns) or singular-plural form agreement (many vs. much), a similar comparative form for multi-syllable adjectives (comparative adjectives), a similar specificity of time (since/for), a similar structure of modal verb + main verb in simple form (modal verbs), and a similar use of the recent past (simple past tense), while the difficult grammar points follow a different use of articles (indefinite article), a different use of the present tense (simple present tense – 3rd person –*s*), a different use of verb + verb combinations (verb complements), and a different use of interrogative forms (*yes/no* questions).

With respect to the finding of learners in Level 9 significantly outperforming learners in Level 5 and that learners in Level 5 performed similarly than learners in Level 7 on the EI test may indicate that implicit mental processes may take longer to settle in the minds of the learners (Robinson, 1997).

A Pearson product moment correlation for the overall scores between the EI test and the MLK test showed a significant moderate positive correlation (r = .31), while a Spearman's rank order correlation between EI and MLK scores by grammar point showed a non-significant negative correlation (rho = -.32, p = .28). Bearing these two correlations in mind, it seems there is a contradiction as indicated in section 5.2, that is, at a global level the scores of the two tests were positively and significantly associated, which might be due to both tests measuring the same grammar points (Absi, 2014; R. Ellis, 2005). The non-significant negative correlation, on the other hand, may be an indication that it is not possible for learners to develop implicit and explicit knowledge of a particular L2 construction at the same time. Hence, there is no such contradiction because a number of learners may find some grammar points implicitly easier to learn and other grammar points explicitly easier to learn.

5.4 The effects of learning difficulty and level groups

Two repeated-measures ANOVAs were conducted to measure the effects of the variables learning difficulty and level groups on the implicit and explicit knowledge scores of the learners. With respect to the learners' explicit knowledge scores, the results showed that there was a significant main effect of learning difficulty with a large effect size indicating that the learners' explicit scores were significantly more accurate on easy grammar points than on difficult grammar points. The results also showed a significant effect for level group with a small effect size indicating that the learners' scores showed greater accuracy at higher levels, and a marginally significant interaction with a small effect size between level group and learning difficulty.

With regard to learners' implicit knowledge scores, a similar pattern of results was found. The results show that there was a significant effect for learning difficulty with a moderate effect size indicating that the learners' scores were significantly more accurate on easy grammar points than on difficult grammar points. Furthermore, the results showed a marginal effect for level group with a small effect size indicating that the learners' implicit scores showed a trend towards greater accuracy at higher levels. The results showed a non-significant interaction between level group and learning difficulty.

The interaction approaching significance for explicit knowledge (MLK) show that there was a tendency for a more-to-less easy vs. difficult contrast from Level 5 to Level 9 as students got more experience with the L2 and more instruction. The non-significant interaction between learning difficulty and level group for the combined implicit scores (EI/ON combined) indicates that it does not matter at which level learners are, there was a tendency for a more-to-less easy vs. difficult contrast from Level 5 to Level 9. Perhaps the interpretation for the marginal significant interaction for explicit knowledge can be explained given the instructional approach learners were instructed with, that is, the more times grammar rules were explained to the learners the more they internalized these in their long-term memory as explicit knowledge, but not as implicit knowledge as evidenced by the absence of an interaction for implicit knowledge. This last argument can be interpreted as the need for more practice of the L2 in real life situations wherein learners can develop implicit knowledge.

In this section the learning difficulty of the targeted grammar points in the current study has been discussed in terms of explicit and implicit knowledge, but other factors such as the IDs in cognitive ability such as language learning aptitude and working memory capacity also have a role to play.

5.5 Learners' language aptitude and working memory

Another finding of the present study involves the relationship between learners' performance on the language learning aptitude test (i.e. LLAMA test) and working memory capacity test (BDS test). The results indicate that the overall language aptitude score (i.e. LLAMA test score; mean % = 50) was higher than the overall working memory score (i.e. BDS test score; mean % = 41). The highest LLAMA score was obtained for LLAMA_E (Sound-symbol correspondence; mean % = 69) and the lowest for LLAMA_D (Sound recognition; mean % = 40). With regard to the language aptitude score by level groups, no significant difference was found across levels. Likewise, no significant difference was found across levels for working memory. When the scores for language aptitude and working memory were correlated, a significant correlation was found (r = .33). As for the principal components analysis that was conducted, all the LLAMA subtests and BDS test loaded on one factor; the loading from the LLAMA_D subtest and the BDS test were weaker than the other subtests. A non-significant correlation was found between the ID variables and the scores of implicit and explicit knowledge, though the association between language aptitude and explicit knowledge approached significance (p = .057). Further analyses by level group revealed that the associations between ID variables and implicit and explicit knowledge, two significant moderate correlations were found between working memory and implicit knowledge in Level 5, and language aptitude and implicit knowledge in Level 7; no significant associations were found between the two ID variables and explicit knowledge.

Although the principal components analysis did not show that LLAMA_D, the sound recognition subtest, loaded separately from the other subtests, as reported in Granena's (2013) study, the correlations between the subtests reveal a similarity with Granena's (2013) study, that is, the pattern of correlations between all pairs of subtests confirm the weak relationship between LLAMA_D and the other subtests, which, on the other hand, are more strongly related to each other. Granena (2013) argues that the subtests LLAMA_B, LLAMA_E, and LLAMA_F could be measuring cognitive abilities involving more explicit cognitive processes such as rote learning, explicit associative learning, and analytical learning. With respect to LLAMA_D, she further argues that this test could be measuring cognitive ability involving more implicit cognitive processes such as memorization of a set of exemplars. In comparison, in the present study all the LLAMA subtests and BDS test loaded on one factor, suggesting no distinction between explicit or implicit abilities. Nevertheless, the loading from the LLAMA_D subtest and the BDS test were weaker than from the other subtests. Furthermore, it is possible that the explicit instruction given to participants in the present study on the LLAMA_D subtest, in which participants were told at the start of the test that they would be tested on whether they had heard certain strings before, could explain the differential loadings from Granena's (2013) study.

Furthermore, participants found the LLAMA_D task the most difficult one (mean % = 40), a result that shares similarity in low score to the BDS test. A similar result was found in Rogers et al.'s (2016) study with regard to the LLAMA_D subtest. These low scores can perhaps be explained, unlike LLAMA_B, LLAMA_E, and LLAMA_F, in the sense that LLAMA_D and the BDS test do not include a study phase that gives learners time to use strategies and problem-solving techniques. In other words, learners did not have the time to employ explicit cognitive processes, but needed to rely to a greater extent on implicit cognitive processes (see Granena, 2013).

Participants scored the highest on the LLAMA_E subtest, both overall (mean % = 69) and at group level (Level 5 = 64%, Level 7 = 75%, Level 9 = 67%). This result is in keeping with Rogers et al.'s (2016) validation study in which participants scored highest on this subtest regardless of gender, L1, formal education qualifications, playing logic puzzles or changing the test timings. A possible explanation why participants in the current study might have found this subtest easier may be that the language learning experience of the learners influenced their performance. The individual group performance in Level 7 was descriptively higher than the other level groups. In Level 7, twenty three out of the thirty learners had studied or were studying one or two foreign languages at the time when the present study took place while in Level 9 there were ten and in Level 5 there were 5.

Returning to the principal components analysis, this analysis shows that the BDS test loaded on the same component as the LLAMA subtests, so it could be said that working memory is a potential component of language aptitude (Miyake & Friedman, 1998; Robinson, 2005; Skehan, 2002). This result is different from that of Roehr and Gánem-Gutiérrez (2009b), in which working memory loaded on a different component than language aptitude.

The difference in results between Roehr and Gánem-Gutiérrez' (2009b) study and the current study is possibly due to the different measures used in each study. In the former, the researchers used the MLAT, a test of L1 reading span, and a test of L2 reading span whereas in the present study the LLAMA test and the BDS test in L1 were employed. The fact that Roehr and Gánem-Gutiérrez (2009b) used two working memory measures and a language aptitude test may explain the difference in results; namely, the MLAT and LLAMA test differ in the inclusion of the factor inductive language learning and the LLAMA_D subtest. This subtest in comparison to the MLAT tasks for phonetic coding ability is more cognitively demanding. For instance, the three tasks in the section of phonetic coding ability on the MLAT test takers identify sounds, connect sounds to graphic symbols, and remember sound-symbols connections, whereas on the LLAMA_D subtest test takers must create phonological representations in working memory. Furthermore, with respect to the WM measures used in each study, Roehr and Gánem-Gutiérrez (2009b) administered two WM measures relying on language material (reading span tests), and the current study employed a WM measure relying on sequences of numbers. In addition to this, the number of participants in each study (39 and 90 respectively) may further explain the difference in results.

Despite the result of the principal components analysis, language aptitude and working memory were treated as two constructs. In the present study, following Sáfár and Kormos' (2008) argument, these two ID variables are related but are not interchangeable considering that the LLAMA and BDS test share approximately 11% of the variance, which is similar to the Standard Hungarian Language Aptitude Test (HUNLAT) and BDS test (13%) in Sáfár and Kormos (2008). The low percentage in the present study may be an indication that the abilities participants used to solve the tasks in working memory and language aptitude were different and at the same time, to some extent, overlapping. The explanation for this lies in the first-order cognitive abilities employed on the BDS test and the higher-order cognitive abilities on the LLAMA test. This means that in order to do the LLAMA test, a combination of first-order and higher-order abilities is required, but no higher-order abilities are required to do the BDS test (See Robinson, 2001, 2002, for a review of first-order and higher-order abilities).

5.6 L2 proficiency, language aptitude and WM as predictors of implicit and explicit knowledge

The result in the present study showed two significant correlations, between working memory and implicit knowledge in Level 5 and language aptitude and implicit knowledge in Level 7. This indicates that the ID variables have a role to play on L2 learning, at least for the implicit knowledge of the targeted grammar points. Bearing these results in mind, it can be argued that both types of cognitive tests may predict L2 learning to some extent. In the understanding that correlations do not show causality, these results need be read in the context of the regression results discussed below.

The results of the regression analyses revealed that the explanatory variables language aptitude and working memory did not significantly predict explicit or implicit knowledge of the targeted difficult and easy grammar points for the cohort of participants as a whole; the only statistical predictor was the control variable L2 proficiency which accounted for 7% of the variance of explicit knowledge of difficult grammar points, 34% of the variance of implicit knowledge of difficult grammar points, and 26% of the variance of implicit knowledge of easy grammar points.

Similarly, the results of the regression analyses for the level groups (Level 5, Level 7, Level 9) of participants indicated that the explanatory variables language aptitude and working memory did not significantly predict explicit knowledge of easy and difficult grammar points once L2 proficiency was controlled for.

In contrast, with respect to the implicit knowledge of difficult grammar points, the results of the regression analysis for the level groups of participants indicated that the explanatory variable language aptitude added the most explanatory power to the model in Level 5, accounting for 12% of the variance, and it approached significance (p = .059). Working memory did not significantly predict implicit knowledge of the targeted difficult grammar points for any group of participants once L2 proficiency was controlled for.

With regard to implicit knowledge of easy grammar points, the results of the regression analysis for the level groups of participants showed that the explanatory variable working memory significantly predicted implicit knowledge of easy grammar points in Level 5, accounting for 12% of the variance. The variable language aptitude did not significantly predict implicit knowledge of the targeted easy grammar points for any group of student participants once L2 proficiency was controlled for.

From the aforementioned summary, the results for the cohort of participants as a whole indicate that L2 proficiency accounts for 7% of the variance in scores on the measure of explicit knowledge of difficult grammar points, 34% of the variance in scores on the measures of implicit knowledge of difficult grammar points, and 26% of the variance in scores on the measures of implicit knowledge of easy grammar points. These results show that the cumulative experience in L2 use is an important factor in L2 learning, and it is probable that "what one already knows is a more important determinant of the knowledge one acquires than one's working memory" (Ackerman, 2007, p. 237) or language aptitude. Put differently, the cumulative experience of the learners in the current study is a variable (Roehr & Gánem-Gutiérrez, 2009b) that influences their performance on the measures of implicit and explicit knowledge of difficult and easy grammar points.

The fact that neither of the ID variables significantly predicted performance on explicit knowledge of difficult or easy grammar points at any of the levels, or for the whole sample of participants, may be explained by the cognitive mechanisms participants may have employed to complete the task, that is, when learners drew on explicit knowledge without time pressure, individual differences in certain cognitive abilities (language aptitude and working memory) are no longer important (though note opposing finding by Granena, 2013b).

Furthermore, the finding that the ID variables aptitude and working memory had no significant influence on the explicit and implicit knowledge of difficult or easy grammar points in the cohort as a whole is driven by the absence of these cognitive variables as predictors in Level 7 and Level 9. This result is possibly due to the teaching approach learners were exposed to, that is, to the deductive instructional approach. Erlam (2005) and Sanz et al. (2014) agree, based on the results of the deductive instruction group in both studies, that students who receive explicit rule explanation and then engage in language production may gain greater control over rhetorical and pragmatic aspects of written or oral discourse, acquisition of more complex grammatical structures and greater knowledge of formal vocabulary. Hence, this finding can explain to some extent that the instruction the participants were exposed to may have levelled out individual differences in cognitive ability.

As for the additional associations between implicit knowledge, explicit knowledge, and the subcomponents of the LLAMA tests for the whole cohort of participants, no significant correlations were found between implicit knowledge and the sub-components of the LLAMA tests. On the other hand, a significant association was found between explicit knowledge and LLAMA_B (Vocabulary learning). This association indicates that there is an interplay between vocabulary learning and explicit knowledge depending on the type of activity learners are engaged in. In other words, the degree of difficulty, the length of exposure, and the degree of novelty of the grammar points (Yalçin and Spada, 2016) may explain, to some extent, such an association.

With respect to the results of the regression analyses between L2 proficiency, sub-components of LLAMA tests and the explicit and implicit knowledge of easy and difficult grammar points for the whole cohort of participants, the findings revealed that vocabulary learning (LLAMA_B) significantly predicted learners' level of explicit knowledge of difficult and easy grammar points, though the amount of variance explained (4% and 5% respectively) was small.

These results are in keeping with Li's (2015) findings that different sub-components of language aptitude show different predictive validity for different aspects of learning, in particular, to explicit and/or implicit knowledge of easy and difficult grammar points. It may be possible that LLAMA_B (Vocabulary learning) was the only significant predictor in learners' level of explicit knowledge of difficult and easy grammar points probably due to the explicit mental mechanisms learners may have developed pertaining to the deductive approach they have experienced as part of their learning process. If the deductive approach involved visual material accompanied by phonetical representations such as pictures of body parts and the pronunciation of each body part, or structures of tenses written on the board followed by the pronunciation of each element in each tense (just to provide a few examples), then it is very likely that learners develop the explicit learning mechanism of association of sounds and symbols. Thus, this ability to learn new words (Carroll and Sapon, 1959; Meara, 2005) and grammar points, for that matter, may have helped learners to develop explicit cognitive and memory processes (Granena, 2013) to learn easy and difficult grammar points. Other studies have found that LLAMA_B contributes to learners' gains on easy grammar points (Yalçin and Spada, 2016) and difficult grammar points (Robinson, 1997). With respect to the lack of prediction of LLAMA_E (Sound-symbol association) and LLAMA_F (Grammatical inferencing) on easy and difficult grammar points, the findings of the current study do not offer an easy interpretation.

The analysis by individual level groups (Level 5, Level 7, and Level 9) demonstrated that only the ID variable language aptitude marginally predicted implicit knowledge of difficult grammar points and working memory significantly predicted implicit knowledge of easy grammar points in less experienced L2 learners (Level 5), but not in more experienced learners (Level 7 and 9) who have had more exposure according to the number of hours of instruction (80 hours per level). In accordance with this, learners in the three level groups differed in terms of reported length of L2 learning and also differed significantly in terms of L2 proficiency. As for the marginal prediction of aptitude, this result suggests that learners may draw on a combination of memory, phonetic coding ability and language analytic ability.

The results in the lower level group (Level 5) in terms of the effect of the explanatory variables language aptitude and complex working memory on implicit knowledge of difficult and easy grammar points are in line with researchers' arguments that aptitude may be a better predictor at lower levels of proficiency if the construct is measured by means of a language aptitude test such as the LLAMA test (Linck et al., 2013; Robinson, 2005; Skehan, 1998).

The same type of argument applies for working memory, which is supported by findings on working memory research where phonological short–term memory tests have been used and the findings suggest that phonological short-term memory is a better predictor at lower levels of proficiency (Hummel, 2009; Juffs & Harrington, 2011). This situation is not parallel to complex working memory research because the findings are less clear. For instance, some researchers using the O-span measure report a greater effect at higher levels (Linck & Weiss, 2011), and others report greater effects at lower levels of proficiency (Serafini & Sanz, 2015). In the present study, complex working memory capacity predicted the successful use of easy grammar points at lower levels (Level 5) but not at higher levels (Level 7 and Level 9). This result is in line with the studies of Coughlin and Tremblay (2013), and Serafini and Sanz (2015). Hence, it seems that complex working memory plays a role with easy grammar points and language analytic ability has no role to play, that is, online processing ability is more important than analysis.

In short, the results in the present study indicate two types of prediction: (1) the cognitive abilities language aptitude and complex working memory predict performance on implicit knowledge measures only, with the exception of the predictive power of LLAMA_B (Vocabulary learning) on explicit knowledge of easy and difficult grammar points, which indicates that learners may have developed explicit mental mechanisms probably due to the prolongued exposure of the deductive approach, and (2) the cognitive abilities language aptitude (marginally) and complex working memory predict performance on the implicit knowledge measures for lower levels of proficiency (Level 5) only. The former finding is consistent with the argument that cognitive abilities are particularly important when learners are “on their own” and working under time pressure (Kormos, 2013; Sanz et al., 2014). Conversely, it appears that when learners are working without time pressure, the cognitive abilities language aptitude and complex working memory do not predict performance once general L2 proficiency is controlled. It appears that learners may draw on memory, phonetic coding ability, and language-analytic ability to significantly predict performance on implicit knowledge of easy and difficult grammar points at lower levels of proficiency only.

5.7 The relationship between background variables and implicit and explicit knowledge

The results on the background variables, namely, use of English at home, use of English at work and attendance at the university self-access centre, revealed a marginal weak negative correlation between implicit knowledge and attendance at the self-access centre, and a significant weak positive correlation between use of English at work and explicit knowledge. For the former association, as argued in the Results chapter, it is possible that the trend towards a negative association indicates that the weaker learners were the ones who attended the self-access centre, or yet another possibility may be that learners attended the self-access centre to fulfil the eight hours as required in the course syllabus, but they did not necessarily work on the language skill or skills they needed to work on to improve their implicit knowledge (feeding into speaking, listening) since their work was not supervised by their class teacher. For the latter association, eight students in Level 9 (see Table 3.2 in section 3.2.2) were enrolled in the English Language Teaching program (E.L.T). In this program learners study the theoretical grounds of teaching and learn teaching techniques and strategies. Given that they informed me that they were teaching English at the same time they were studying in the E.L.T. program when the research took place, it is plausible that this working experience may have reinforced their explicit grammar skills (Celce-Murcia & Larsen-Freeman, 1999).

CHAPTER SIX: CONCLUSION

The present study was aimed at examining the relationship between implicit and explicit knowledge and the relationship of each type of knowledge with language learning aptitude and working memory capacity in Mexican learners of L2 English at three different levels of proficiency (intermediate, upper-intermediate, advanced). Explicit knowledge was assessed through a MLK test comprising error correction, rule explanation, and rule illustration tasks. On the other hand, implicit knowledge was measured by means of an oral elicited imitation test and an oral narrative test. With respect to language learning aptitude and working memory, the former was operationalised by the LLAMA test, and the latter by the BDS test. The current study was also an investigation of the relationship between teachers' and learners' perceived difficulty of grammar points, and learners' actual explicit and implicit performance on these measures. Furthermore, comparisons were made between learners' scores on the implicit and explicit measure as well as on the cognitive measures in the three different levels of proficiency.

6.1 Summary of findings

As a first step to inform subsequent analyses of relationships between implicit and explicit knowledge, the present study asked the student participants and the teacher participants to judge the learning difficulty of the targeted grammar points using a 5-point scale. I used their judgements to categorize each grammar point into easy or difficult. From a methodological perspective, the current study demonstrates the advantages of categorizing grammar points following this procedure, which increases the validity of classifying grammar points into easy and difficult based on learners' and teachers' judgments under the same teaching and learning L2 context. This type of procedure has proved to be useful in other studies (Absi, 2014; Rodríguez Silva & Roehr-Brackin, 2016).

The finding of the similarity of learners' overall performance on implicit and explicit measures suggests that in form-focused teaching approaches learners may develop both implicit and explicit knowledge of grammar points to the same extent, but this finding should be taken with caution considering that not all grammar points were covered in both oral tasks (i.e. EI and ON test) as in the written task (i.e. MLK test) in which Part 1 and 2 of the test cover the 13 targeted grammar points. Nevertheless, it is essential to be aware that learners instructed with a traditional deductive approach such as the presentation-practice-production approach learners in the present study were exposed to, implicit and explicit gains are obtained.

In terms of the relationship between learners' overall implicit and explicit knowledge, the positive and weak correlation indicates that learners develop both types of knowledge to some extent. This was further evidenced by the analysis by grammar point. As discussed in the previous chapter, the negative correlation by grammar point indicates that learners' implicit and explicit knowledge diverge significantly in the highest level group (Level 9), but not in the other levels.

With regard to the relationship between learners' implicit and explicit knowledge with language learning aptitude and working memory by level groups, the findings indicate that the cognitive variables play a role in implicit L2 learning but not in explicit learning in Level 5 only.

In terms of the regression analyses for implicit and explicit knowledge, the findings suggest that at lower levels of proficiency (Level 5) learners draw on a combination of memory, phonetic coding ability and language-analytic ability to complete implicit knowledge tasks. This result informs of the relevance of IDs in cognitive abilities in lower levels of proficiency as reliable predictors of learners' ability to use grammar points in communicative tasks. Thus, the finding supports researchers' (Linck et al., 2013; Robinson, 2005; Serafini & Sanz, 2015) argument that language aptitude and working memory are better predictors of L2 learning at lower levels of proficiency.

To sum up, these results indicate that at higher levels of proficiency and with more cumulative time in the L2 classroom, language aptitude and complex working memory cease to play a role as predictors of implicit and explicit knowledge, arguably due to the deductive approach in the presentation-practice-production tradition the learners in the present study were exposed to. In other words, as learners progress and their proficiency increases, this deductive approach appears to eliminate the role of individual differences in cognitive abilities (language aptitude and working memory), thus serving as an equalizer. This finding is consistent with the results in Erlam's (2006) study, in which aptitude and/or working memory were important in the context of two experimental instructional approaches (inductive and structured input-based), but not in the context of a third, more conventional approach (deductive) that the participants were used to from their regular classes.

6.2 Pedagogical implications

From a pedagogical perspective, the findings of the present study provide several implications that are useful for L2 teachers using a deductive instructional approach. One interesting finding in the present study is the fact that grammar points can be classified as easy and difficult drawing on the judgements of teachers and learners (Absi, 2014; Rodríguez Silva & Roehr-Brackin, 2016). The categorization of grammar points may facilitate teachers the organization of activities and the allocation of time for each activity depending on the type of grammatical structure to teach.

According to the results in the current study, learners' performance on explicit and implicit measures show that they struggled more with difficult grammar points than easy grammar points for both types of measures. This is an indication that teachers should be aware that some grammar points (e.g. easy grammar points) may take less time for learners to learn both explicitly and implicitly, and it could be argued that it would take teachers less time to teach them. However, teachers should also be aware that other grammar points (e.g. difficult grammar points) may take learners more time to learn them both explicitly and implicitly (DeKeyser, 2005; Ellis, 2006; Thepseenu and Roehr, 2013; Ziętek and Roehr, 2011). It may be too cumbersome for teachers to consider the learning difficulty of grammar points, but this is an important element in the learning process of their learners and they should not disregard this factor.

The finding of the negative correlation between explicit and implicit knowledge by grammar point suggests that teachers should take learners' development of implicit and explicit knowledge of easy and difficult grammar points into consideration for lesson planning given that learners may not be able to develop both types of knowledge at the same time. In other words, teachers need to understand that some structures may be more difficult to learn as explicit knowledge and others as implicit knowledge, and based on this, teachers need to select activities accordingly.

Another pedagogical implication of the present study is that teachers should be aware of the relevance of the IDs of learners with regard to language aptitude and working memory as shown by the significant associations between working memory and implicit knowledge in Level 5, and language aptitude and implicit knowledge in Level 7. It is important for teachers to be aware that learners use these cognitive abilities to learn the L2, and it is the teachers' responsibility to make learners aware that they can use such abilities in their favour to overcome learning obstacles such as learning difficult grammar points. By taking into account these cognitive abilities, teachers can facilitate the process of L2 learning to learners who have more difficulty with the language by either diagnosing deficiencies in vocabulary learning, identification of sounds, or deducing the underlying rules of the L2, and by attending this situation, they would be avoiding treating learners as average speakers or listeners without paying much attention to individual differences (Skehan, 2012; Roberts & Meyer, 2012). Hence, these results in the current study informed that learners not only differ in their learning capacity, but also in the stage of development of their implicit and explicit knowledge according to their levels of proficiency.

The finding that the ID variable language aptitude had an influence on the implicit knowledge of difficult grammar points and working memory on implicit knowledge of easy grammar points in less experienced L2 learners (Level 5) suggests the pedagogical implication that by testing learners on these two cognitive abilities, and knowing what strengths and limitations in working memory and language aptitude learners have, teachers can provide further support to their L2 learning (Roberts & Meyer, 2012) either by focusing on phonological problems or vocabulary learning.

DeKeyser (2012) and Skehan (2012) seem to agree that it is up to teachers to pay attention to the cognitive abilities of language aptitude and working memory, but such a task could be overwhelming considering that teachers have to cover a course syllabus and design teaching material. Teachers can pay attention to these two cognitive abilities by observing which learners are not efficient in learning grammar, learning vocabulary, or pronouncing words (i.e. language aptitude) as well as observing which learners are not able to follow instructions or retain the main key concepts of a reading (i.e. working memory). By doing this, teachers can identify learners with low or high languge aptitude and/or working memory. Following on this, teachers may be able to support those learners with low language aptitude and/or working memory, so these learners can learn the L2 as efficiently as those learners with high language aptitude and/or working memory; even better, learners with low or high language aptitude and/or working memory can be matched with specific learning activities or with appropriate methodologies (Wesche, 1981) if possible.

6.3 Limitations of the study

A number of limitations of this study are worth mentioning. The first one has to do with the number of grammar points included on the ON test. Recall that the same 13 grammar points were targeted on the MLK test (explicit measure) and the EI test (implicit measure). A combined implicit score (EI/ON combined scores) was adopted to have a fuller representation of learners' implicit knowledge. Originally, 8 targeted grammar points were included on the ON test, but learners only produced the minimum required number of suppliances (3 occurrences) for 5 targeted grammar points. If the same number of targeted grammar points were included on the ON test, perhaps a different combined implicit score would result.

One more limitation of the current study is the lack of implementation of think-aloud protocols to see why teachers and learners judged some grammar points easy and others difficult. This procedure would allow clarifying why learners judged most grammar points as easier and teachers as more difficult. Following on the lack of think-aloud protocols in the present study, another limitation is the lack of classroom observations to find out to what extent the 26 participating teachers used a deductive approach.

Likewise, it would be useful to know whether teachers categorized each grammar point as easy or difficult because they were either assessing them in terms of the learning difficulty they saw in their learners, or the learning difficulty they themselves had with these grammar points when they were learning the language, or the teaching difficulty they found for each grammar point, or the combination of all these issues.

6.4 Suggestions for further research

To my knowledge, the current study is the first empirical attempt where learning difficulty of 13 grammar points is related to learners' performance on implicit and explicit knowledge measures, and their performance on language learning aptitude and working memory capacity in three different levels of proficiency (intermediate, upper-intermediate, advanced).

One suggestion for further research, as noted earlier, the few grammar points included on the oral narrative test should be increased, or better yet, have the same number of grammar points on both implicit measures to obtain a broader representation of learners' implicit knowledge. Perhaps a better choice is to include one or two more oral narrative tests to include all targeted grammar points considering that including 13 easy and difficult grammar points in one narrative task will not yield enough target grammar points in obligatory contexts, as it was the case in the present study. Adopting this type of procedure would provide researchers with an overall implicit score of learners' performance on all targeted grammar points, as they did on the EI test.

The current study aimed at examining learners' performance at intermediate (Level 5), upper-intermediate (Level 7), and advanced (Level 9). Further research with lower levels of proficiency (e.g. level 3 and level 4) in a similar L2 learning context may yield different results. On one hand, it would be interesting to see whether learners' judgements of targeted grammar points would be similar or different in comparison to the judgements obtained in the present study. On the other hand, it would also be interesting to examine whether they would perform better on the MLK test considering that lower-level learners may focus more on explicit learning of the language. In the same vein, further research with lower levels of proficiency in a similar learning context may yield different results with regard to the role of aptitude and working memory in such lower-level learners. Studies including lower-level learners would allow clarifying whether the IDs in language aptitude and working memory would be even better predictors at these levels, and whether they would also predict explicit knowledge in beginner learners.

Another fruitful investigation would be conducting research in communicative and immersive settings. In these types of settings learners are more exposed to naturalistic input and communicative activities (Absi, 2014) and it would be interesting to explore learners' performance on both implicit and explicit measures. Assuming that such learners are likely to have better implicit knowledge, it would be of interest to see to what extent the effect of this type of instruction may influence their explicit learning of grammatical constructions. In other words, investigation on this type of settings might inform whether inductive learning can have an effect on learners' understanding of grammatical rules. By the same token, it would be of interest to investigate whether the role of the IDs in language aptitude and working memory would be levelled out after some time in meaning-focused and/or immersion settings.

In conclusion, this study of learners' and teachers' perceptions of learning difficulty of grammar points in relation to their L2 proficiency, implicit and explicit knowledge, language aptitude, and working memory, presents a potentially productive area of research. The results of the present study provide useful information for L2 researchers, L2 teachers, and L2 learners in the sense that teachers can directly ask their learners what structures they find easy and difficult to learn and pay more attention in the instruction of those grammar points. Teachers can also test learners on their language learning aptitude and working memory capacity and further support those learners with deficiencies in either one of these two cognitive abilities or both of them to improve their L2 learning process.

REFERENCES

Abrahamsson, N., & Hyltenstam, K. (2008). The robustness of aptitude effects in near-native second language acquisition. *Studies in Second Language Acquisition*, *30*, 481-509.

Absi, Z. (2014). *Metalinguistic knowledge and speaking proficiency in Syrian university-level learners of English* (Unpublished doctoral thesis). University of Essex.

Ackerman, P. L. (2007). New developments in understanding skilled performance. *Current Directions in Psychological Research*, *16*, 235-239.

Akakura, M. (2012). Evaluating the effectiveness of explicit instruction on implicit and explicit L2 knowledge. *Language Teaching Research, 16*(1), 9-37.

Alderson, J. C. Clapham, C., & Steel, D. (1997). Metalinguistic knowledge, language aptitude and language proficiency. *Language Teaching Research*, *1*, 93-121.

Alipour, S. (2014). Metalinguistic and linguistic knowledge in foreign language learners. *Theory and Practice in Language Studies, 4*(12), 2640-2645.

Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review, 89*, 369-406.

Anderson, J. R. (1996). A simple theory of complex cognition. *American Psychologist, 51*(4), 355-365.

Anderson, J. R. (2005). *Cognitive psychology and its implications* (6th ed.). New York, NY: Worth Publishers.

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417-423.

Baddeley, A. D. (2003). Working memory and language: An overview. *Journal of Communication Disorders*, *36*, 189-209.

Baddeley, A. D. (2003b). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, *4*, 829-839.

Baddeley, A. (2007). *Working memory, thought, and action.* Oxford: Oxford University Press.

Baddeley, A. D. (2012). Working memory, theories, models, and controversies. *Annual Review of Psychology*, *63*, 1-29.

Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation.* (pp. 47-89). New York: Academic Press.

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multicomponent model. In A. Miyake and P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). Cambridge: Cambridge University Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behaviour, 14,* 575-589.

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning, 28*(1), 69-83.

Bialystok, E. (1979). Explicit and implicit judgements of L2 grammaticality. *Language Learning, 29*, 81-103.

Bialystok, E. (1994). Analysis and control in the development of second language proficiency. *Studies in Second Language Acquisition, 16,* 157-168.

Biber, D., Conrad, S., & Leech, G. (2002). *Longman student grammar of spoken and written English.* Harlow: Longman.

Bowles, M. (2011). Measuring implicit and explicit linguistic knowledge: What can heritage language learners contribute? *Studies in Second Language Acquisition, 33,* 247-271.

Brown, R. (1973). *A first language: The early stages.* Cambridge, MA: Harvard University Press.

Bulté, B., & Housen, A. (2012). Defining and operationalising L2 complexity. In A. Housen, F. Kuiken, & I. Vedder (Eds.), *Dimensions of L2 performance and proficiency* (pp. 21-46). Amsterdam: John Benjamins.

Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2010). The role of language aptitude in first language attrition: The case of pre-pubescent attriters. *Applied Linguistics 31*(3), 443-464.

Carroll, J. B. (1981). Twenty-five years of research on foreign language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude* (pp. 83-118). Rowley, MA: Newbury House.

Carroll, & Sapon (1959). *Modern language aptitude test*. New York: The Psychological Corporation/Harcourt Brace Jovanovich.

Castañeda-Jiménez, G., & Jarvis, S. (2013). Exploring lexical diversity in second language Spanish. In K. Geeslin (Ed.), *The handbook of Spanish second language acquisition* (pp. 498-513). Boston: Wiley.

Cherciov, M. (2011). *Between attrition and acquisition: The dynamics between two languages in adult migrants* (Unpublished doctoral dissertation). University of Toronto.

Cheung, H. (1996). Nonword span as a unique predictor of second-language vocabulary learning. *Developmental Psychology, 32*(5), 867-873.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course.* Boston, MA: Heinle and Heinle.

Cerezo, L., Caracas, A., & Leow, R. (2016). The effectiveness of guided induction versus deductive instruction on the development of complex Spanish *gustar* structures: An analysis of learning outcomes and processes. *Studies in Second Language Acquisition, 38*, 265-291.

Cohen, J. (1988). *Statistical power analysis for the behavioural sciences* (2nd ed.). New York: Academic Press.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). New York: Routledge.

Collins, L., Trofimovich, P., White, J., Cardoso, W., & Horst, M. (2009). Some input on the easy/difficult grammar question: An empirical study. *Modern Language Journal, 93*(3), 336-353.

Colom, R., Shih, P. C. Flores-Mendoza, C., & Quiroga, M. A. (2006). The real relationship between short-term memory and working memory. *Memory 14*, 804-813.

Cook, V. J. (1996). *Second language learning and language teaching* (2nd ed.). London: Edward Arnold.

Cook, V. J. (2001). *Second language learning and language teaching* (3rd ed.). London: Edward Arnold.

Coughlin, C. E., & Tremblay, A. (2013). Proficiency and working memory based explanations for non-native speakers' sensitivity to agreement in sentence processing. *Applied Psycholinguistics, 34*, 615-646.

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism, 19*, 121-129.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behaviour*, *19*(4), 450-466.

De Graaff, R. (1997). The eXperanto experiment: Effects of explicit instruction on second language acquisition. *Studies in Second Language Acquisition*, *19,* 249-297.

DeKeyser, R. M. (1998). Beyond focus on form: Cognitive perspectives on learning and practicing second language grammar. In C. Doughty, & J. Williams (Eds.), *Focus on form in classroom second language acquisition* (pp. 42-63). Cambridge, England: Cambridge University Press.

DeKeyser, R. M. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499-533.

DeKeyser, R. M. (2003). Implicit and explicit learning. In C. J. Doughty, & M. H. Long, (Eds.). *The handbook of second language acquisition*. Oxford: Blackwell.

DeKeyser, R. M. (2005). What makes learning second-language grammar difficult? A review of issues. *Language Learning*, *55*(S1), 1-25.

DeKeyser, R. M. (2007). Skill acquisition theory. In B. VanPatten, & J. Williams (Eds.), *Theories in second language acquisition: An introduction* (pp. 91-113). Mahwah, NJ: Erlbaum.

DeKeyser, R. M. (2012). Interactions between individual differences, treatments, and structures in SLA. *Language Learning, 62* (S2), 189-200.

DeKeyser, R. M. (2016). Of moving targets and chameleons. *Studies in Second Language Acquisition, 38,* 353-363.

Della Putta, P. (2016). The effects of textual enhancement on the acquisition of two nonparallel grammatical features by Spanish-speaking learners of Italian. *Studies in Second Language Acquisition, 38,* 217-238.

Dietz, G. (2002). On rule complexity. A structural approach. *EUROSLA Yearbook, 2,* 263-286.

Dixson, R. J. & Andújar, J. I. (1967). *Resúmen práctico de la gramática Inglesa.* Englewood, NJ: Prentice Hall.

Dörnyei, Z. (2005). *The psychology of the language learner.* London: Lawrence Erlbaum.

Dörnyei, Z. (2009). *The psychology of second language acquisition*. Oxford: Oxford University Press.

Douglas, D. (2001). Performance consistency in second language acquisition and language testing research: A conceptual gap. *Second Language Research, 17*(4), 442-456.

Eastwood, J. (1999). *Oxford Practice Grammar with answers* (New Edition). Oxford: Oxford University Press.

Ehrman, M. E. & Oxford, R. L. (1995). Cognition plus: Correlates of language learning success. *Modern Language Journal, 79,* 67-89.

Elder, C., Warren, J., Hajek, J., Manwaring, D., & Davies, A. (1999). Metalinguistic knowledge: How important is it in studying a language at university? *Australian Review of Applied Linguistics, 22*, 81-95.

Ellis, N. C. (1994). Implicit and explicit language learning: An overview. *Implicit and explicit language learning of languages*. Academic Press.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18*, 91-126

Ellis, N. C. (2005). At the interface: Dynamic interactions of explicit and implicit language knowledge. *Studies in Second Language Acquisition, 27,* 305-352.

Ellis, N. C. (2006a). Language acquisition as rational contingency learning. *Applied Linguistics, 27*, 1-24.

Ellis, N. C. (2006b). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics, 27*(2), 164-194.

Ellis, N. C. (2011). Implicit and explicit SLA and their interface. In C. Sanz and R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 35-47). Washington, D.C.: Georgetown University Press.

Ellis, N. C. & Larsen-Freeman, D. (2006). Language emergence: Implications for applied linguistics. *Applied Linguistics, 27*(4), 558-589.

Ellis, R. (1990). *Instructed second language acquisition.* Oxford, UK: Blackwell.

Ellis, R. (1993). The structural syllabus and second language acquisition. *TESOL Quaterly, 27*, 91-113.

Ellis, R. (2004). The definition and measurement of L2 explicit knowledge. *Language Learning, 54*, 227-275.

Ellis, R. (2005). Measuring implicit and explicit knowledge of a second language: A psychometric study. *Studies in Second Language Acquisition, 27*(2), 141-72.

Ellis, R. (2006). Modelling learning difficulty and second language proficiency: The differential contributions of implicit and explicit knowledge. *Applied Linguistics, 27*(3), 431-463.

Ellis, R., & Barkhuizen, G. (2005). *Analysing learner language.* Oxford: Oxford University Press.

Ellis, R. & Loewen, S. (2007). Confirming the operational definitions of explicit and implicit knowledge in Ellis (2005): Responding to Isemonger. *Studies in Second Language Acquisition, 29*, 119-126.

Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J., & Reinders, H. (2009). *Implicit and explicit knowledge in second language learning.* Bristol: Multilingual Matters.

Erçetin, G., & Alptekin, C. (2013). The implicit/explicit knowledge distinction and working memory: Implications for second-language reading comprehension. *Applied Psycholinguistics*, *34*(4), 727-753.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *The Psychological Review, 102*(2), 211-245.

Erlam, R. (2005). Language aptitude and its relationship to instructional effectiveness in second language acquisition. *Language Teaching Research*, *9*(2), 147-171.

Erlam, R. (2006). Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study. *Applied Linguistics, 27*(3), 464-491.

Forsberg, F., & Sandgren, M. (2013). High-level proficiency in late L2 acquisition: Relationships between collocational production, language aptitude and personality. In G. Granena, & M. Long (Eds.). *Sensitive periods, language aptitude, and ultimate L2 attainment.* (pp. 231-255). Amsterdam: John Benjamins.

Fotos, S. (1993). Consciousness raising and noticing through focus on form: Grammar task performance versus formal instruction. *Applied Linguistics, 14*(4), 385-407.

Gass, S., & Selinker, L. (2008). *Second language acquisition: An introductory course* (3rd edition). New York: Routledge.

Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences, 3*(11), 410-419.

Gathercole, S. E., & Baddeley. A. D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language, 28*(2), 200-213.

Gilabert, R., & Muñoz, C. (2010). Differences in attainment and performance in a foreign language: The role of working memory capacity. *International Journal of English Studies, 10*(1), 19-42.

Goldschneider, J., & DeKeyser, R. (2001). Explaining the “natural order of L2 morpheme acquisition” in English: A meta-analysis of multiple determinants. *Language Learning, 51*, 1-50.

Goo, J., Granena, G. Yilmaz, Y., & Novella, M. (2015). Implicit and explicit instruction in L2 learning: Norris and Ortega (2000) revisited and updated. In P. Rebuschat (Ed.), *Implicit and explicit learning of languages* (pp. 443-482). Amsterdam: John Benjamins.

Granena, G. (2012). *Age differences and cognitive aptitudes for implicit and explicit learning in ultimate L2 attainment* (Unpublished doctoral dissertation). University of Maryland.

Granena, G. (2013a). Cognitive aptitudes for second language learning and the LLAMA language aptitude test. In G. Granena & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 105-129). Amsterdam: John Benjamins.

Granena, G. (2013b). Reexamining the robustness of aptitude in second language acquisition. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment* (pp. 179-204). Amsterdam: John Benjamins.

Granena, G. (2014). Language aptitude and long-term achievement in early childhood L2 learners. *Applied Linguistics, 35*(4), 483-503.

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research*, *29*(3), 311-343.

Graus, J., & Coppen, P. A. (2015). Defining grammatical difficulty: A student teacher perspective. *Language Awareness, 24*(2), 101-122.

Green, P. S., & Hetch, K. (1992). Implicit and explicit grammar: An empirical study. *Applied Linguistics*, *13*(2), 168-184.

Grigorenko, E. L., Sternberg, R. J., & Ehrman, M. (2002). A theory based approach to the measurement of foreign language learning ability: The Canal-F theory and test. *The Modern Language Journal*, *84*(3), 390-405.

Gutiérrez, X. (2013). The construct validity of grammaticality judgement tests as measures of implicit and explicit knowledge. *Studies in Second Language Acquisition,* 1-27.

Hale, J. B., Hoeppner, J. B., & Fiorello, C. A. (2002). Analyzing digit span components for assessment of attentinal processes. *Journal of Psychoeducational Assessment, 20,* 128-143.

Han, Y., & Ellis, R. (1998). Implicit knowledge, explicit knowledge and general language proficiency. *Language Teaching Research, 2*(1), 1-23.

Harley, B., & Hart, D. (1997). Language aptitude and second language proficiency in classroom learners of different starting ages. *Studies in Second Language Acquisition, 19,* 379-400.

Harley, B., & Hart, D. (2002). Age, aptitude, and second language learning on a bilingual exchange. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 301-330). Philadelphia: Benjamins.

Harrington, M. W., & Sawyer, M. (1992). L2 working memory capacity and L2 reading skills. *Studies in Second Language Acquisition*, *14*(1), 25-38.

Housen, A., Pierrard, M., & Van Daele, S. (2005). Structure complexity and the efficacy of explicit grammar instruction. *The Modern Language Journal, 91*(3), 235-269.

Housen, A., & Simeons, H. (2016). Introduction to special issue. *Studies in Second Language Acquisition*, *38*(2), 163-175.

Hu, G. (2002). Psychological constraints on the utility of metalinguistic knowledge in second language production. *Studies in Second Language Acquisition, 27(2),* 347-86.

Huang, W. (2012). *Learning difficulty of pedagogical grammar in L2 English: theory, teachers' and learners' judgement* (Unpublished MA dissertation). University of Essex.

Hulstijn, J. H. (2002). Towards a unified account of the representation, processing and acquisition of second language knowledge. *Second Language Research, 18*(3), 193-223.

Hulstijn, J. H. (2005). Theoretical and empirical issues in the study of implicit and explicit second-language learning: Introduction. *Studies in Second Language Acquisition, 27*(2), 129-140.

Hulstijn, J. H., & De Graaff, R. (1994). Under what conditions does explicit knowledge of a second language facilitate the acquisition of implicit knowledge? A research proposal. *AILA Review, 11*, 97-112.

Hummel, K. M. (2009). Aptitude, phonological memory, and second language proficiency in nonnovice adult learners. *Applied Psycholinguistics, 30*, 225-249.

Isemonger, I. M. (2007). Operational definitions of implicit and explicit knowledge: Response to R. Ellis (2005) and some recommendations for future research in this area. *Studies in Second Language Acquisition, 29,* 101-118.

Iwashita, N., Brown, A., McNamara, T., & O'Hagan, S. (2008). Assessed levels of second language speaking proficiency: How distinct? *Applied Linguistics, 29*(1), 24-49.

Juffs, A. (2004). Representation, processing and working memory in a second language. *Transactions of the Philological Society, 102*(2), 119-225.

Just, M. A., & Carpenter, P. (1992). A capacity theory of comprehension: Individual differences in working memory. *The Psychological Review, 99*, 122-149.

Juffs, A., & Harrington, M. (2011). Aspects of working memory in L2 learning. *Language Teaching, 44*(2), 137-166.

Kay, S., & Jones, V. (2003). *American Inside Out. Intermediate Student´s Book.* Oxford: McMillan.

Kay, S., Jones, V., & Gomm, H. (2009). *New American Inside Out. Pre-Intermediate Student´s Book.* Oxford: McMillan.

Kim, Y., Jung, Y., & Tracey-Ventura, N. (2016). A measure of proficiency or short-term memory? Validation of an elicited imitation test for SLA research. *The Modern Language Journal, 100*(3), 655-673.

Kormos, J. (2013). New conceptualizations of language aptitude in second language attainment. In G. Granena, & M. Long (Eds.), *Sensitive periods, language aptitude, and ultimate L2 attainment.* Amsterdam, NLD: John Benjamins Publishing Company.

Kormos, J., & Sáfár, A. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language and Cognition, 11*(2), 261-271.

Krashen, S. D. (1981). *Second language acquisition and second language learning*. Oxford: Pergamon.

Krashen, S. D. (1982). *Principles and practice in second language acquisition.* Englewood Cliffs, NJ: PrenticeHall.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. London: Longman.

Krashen, S. D. (1992). Under what circumstances, if any, should formal grammar instruction take place? *TESOL Quaterly, 26,* 409-411.

Krashen, S. D. (1994). The input hypothesis and its rivals. In N. Ellis (Ed.), *Implicit and explicit learning of language* (pp. 45-77). London: Academic Press.

Larsen-Freeman, D., & Long, M. H. (1991). *Introduction to second language acquisition research.* London: Longman.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. Routledge: New York.

Leclercq, P., & Edmonds, A. (2014). How to assess L2 proficiency? An overview of proficiency assessment research. In P. Leclercq, A. Edmonds, & Hilton, H. (Eds.), *Measuring L2 proficiency: Perspectives from SLA*, (pp. 3-23). Multilingual Matters.

Li, S. (2014). The associations between language aptitude and second language grammar acquisition: A meta-analytic review of five decades of research. *Applied Linguistics, 36*(3), 385-408.

Li, S. (2015). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition, 38*(2), 1-40.

Lichtman, K. (2013). Developmental comparisons of implicit and explicit language learning. *Language Acquisition, 20*(2), 93-108.

Lightbown, P. M., & Spada, N. (2001). *How languages are learned.* Oxford: Oxford University Press.

Linck, J. A., & Weiss, D. J. (2011). Working memory predicts the acquisition of explicit L2 knowledge. In C. Sanz, & R. P. Leow (Eds.), *Implicit and explicit language learning: Conditions, processes, and knowledge in SLA and bilingualism* (pp. 101-113). Washington, DC: Georgetown University Press.

Linck, J. A., Hughes, M. M., Campbell, S. G., Silbert, N. H., Tare, M., Jackson, S. R.,…Doughty, C. J. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning, 63*(3), 530-566.

Long, M. H., & Doughty, C. J. (2009). *Handbook of language teaching.* Oxford: Blackwell.

Loewen, S. (2009). Grammaticality judgement tests and the measurement of implicit and explicit L2 knowledge. In R. Ellis, S. Loewen, C. Elder, R. Erlam, J. Philp, & H. Reinders (Eds.), *Implicit and explicit knowledge in second language learning, testing and teaching* (pp. 94-112). Tonawanda, NY: Multilingual Matters.

Meara, P. (2005). *LLAMA language aptitude tests*: *The Manual*. Lognostics: University of Swansea, (www.lognostics.co.uk/tools/llama).

Meara, P., Milton, J., & Lorenzo-Dus, N. (2001). *Language Aptitude Tests manual*. Express Publishing.

Mitchell, A. E., Jarvis, S., O'Malley, M., & Konstantinova, I. (2015). Working memory measures and L2 proficiency. In Z. Wen, M. Borges & A. McNeill (Eds.),*Working memory in second language acquisition and processing* (pp. 270-283). United Kingdom: Multilingual Matters.

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Workin memory as language aptitude. In A. F. Healy and L. E. Bourne (Eds.), *Foreign language learning: Psycholinguistic studies on retention and training* (pp. 339-364).. Mahwah, NJ: Lawrence Erlbaum.

Mizera, G. J. (2006). *Working memory and L2 oral* fluency (Unpublished Ph.D. dissertation). University of Pittsburgh.

Murphy, R. (2006). *English grammar in use* (3rd ed.). Cambridge: Cambridge University Press.

Norman, G. (2010). Likert scales, levels of measurements and the "laws" of statistics. *Avances in Health Sciences Education: Theory and Practice*, *15*(5), 625-632.

Norris, J. M., & Ortega, L. (2001). Does type of instruction make a difference? Substantive findings from a meta-analysis review. *Language Learning*, *51*(1), 157-213.

Ortega, L. (2009). Foreing language aptitude. In L. Ortega, *Understanding second language acquisition* (pp. 145-167). London: Hodder Education.

Osaka, M., & Osaka, N. (1992). Language independent working memory as measured by Japanese and English reading span tests. *Bulletin of the Psychonomic Society, 30*, 287-289.

Pallant, J. (2010). *SPSS survival manual* (4th edition). Maidenhead: Open University Press.

Palloti, G. (2015). A simple view of linguistic complexity. *Second Language Research*, *31*(1), 117-134.

Paradis, M. (1994). Neurolinguistic aspects of implicit and explicit memory: Implications for bilingualism. In N. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 393-419). London: Academic Press.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism.* Amsterdam: John Benjamins.

Paradis, M. (2009). *Declarative and procedural determinants of second languages*. Amsterdam: John Benjamins.

Petersen, C. R., & Al-Haik, A. (1976). The development of the Defense Language Aptitude Battery (DLAB). *Educational and Psychological Measurement*, *36*, 369-380.

Pienemann, M. (1999). *Language processing and second language development: Processability Theory*. Amsterdam: John Benjamins.

Pimsleur, P. (1966). Testing foreign language learning. In Albert Valdman (Ed.), *Trends in language teaching*. New York: McGraw-Hill Company.

Ranta, L. (2002). The role of learners' language analytic ability in the communicative classroom. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 159-180). Amsterdam/Philadelphia: John Benjamins.

Ranta, L. (2008). Aptitude and good language learners. In C. Griffiths (Ed.) *Lessons from good language learners*. Cambridge: Cambridge University Press.

Reber, A. (1976). Implicit learning of synthetic languages: The role of instructional set. *Journal of Experimental Psychology, 2*(1), 88-94.

Reber, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, *118*, 219-235.

Reber, A. (1993). *Implicit learning and tacit knowledge*. Oxford: Clarendon.

Rebuschat, P. (2013). Measuring implicit and explicit knowledge in second language research. *Language Learning, 63*(3), 595-626.

Rebuschat, P., & Williams, J. N. (2012). Implicit and explicit knowledge in second language acquisition. *Applied Psycholinguistics, 33,* 829-856.

Reingold, E. M., & Merikle, P. M. (1988). Using direct and indirect measures to study perception without awareness. *Perception and Psychophysics*, *44*, 563-375.

Reves, T. (1982). *What makes a good language learner?* (Unpublished PhD thesis). Hebrew University of Jerusalem.

Révész, A., Michel, M., & Gilabert, R. (2016). Measuring cognitive task demands using dual-task methodology, subjective self-ratings, and expert judgements: A validation study. *Studies in Second Language Acquisition*, doi: 10.1017/S0272263115000339, 1-35.

Richards, J., & Rogers, T. (1986). *Approaches and methods in language teaching.* Cambridge: Cambridge University Press.

Richards, J. C., & Schmidt, R. (2002). *Dictionary of language teaching and applied linguistics* (3rd ed.). London, England: Longman.

Roberts, L., & Meyer, A. (2012). Individual differences in second language learning: Introduction. *Language Learning, 62*(Suppl. 2), 1-4.

Robinson, P. (1996). Learning simple and complex second language rules under implicit, incidental, rule-search and instructed conditions. *Studies in Second Language Acquisition, 18*, 27-67.

Robinson, P. (1997). Individual differences and the fundamental similarity of implicit and explicit adult L2 learning. *Language Learning, 46*, 45-99.

Robinson, P. (2001). Individual differences, cognitive abilities, aptitude complexes, and learning conditions in SLA. *Second Language Research, 17*(4), 368-392.

Robinson, P. (2002). *Individual differences and instructed language learning*. Amsterdam: John Benjamins.

Robinson, P. (2005). Aptitude and second language acquisition. *Annual Review of Applied Linguistics, 25*, 46-73.

Robinson, P. (2007). Aptitudes, abilities, contexts and practice. In R. M. DeKeyser (Ed.), *Practice in second language learning: Perspectives from applied linguistics and cognitive psychology* (pp. 256-286). Cambridge: Cambridge University Press.

Rodríguez Silva, L. H. (2013). *Learning difficulty, explicit and implicit knowledge of L2 English grammar: A study with L1 Spanish adult learners in* Mexico (Unpublished MA Dissertation). University of Essex.

Rodríguez Silva, L. H., & Roehr-Brackin, K. (2016). Perceived learning difficulty and actual performance: Explicit and implicit knowledge of L2 English grammar points among instructed adult learners. *Studies in Second Language Acquisition, 38*(2), 317-340.

Roehr, K. (2005). *Metalinguistic knowledge in second language learning: An emergentist perspective* (Unpublished Ph.D. Thesis). University of Lancaster.

Roehr, K. (2008). Metalinguistic knowledge and language ability in university-level L2 learners. *Applied Linguistics, 29*(2), 173-199.

Roehr, K., & Gánem-Gutiérrez, G. A. (2009a). Metalinguistic knowledge: A stepping stone towards L2 proficiency. In A. Benati (Ed.), *Issues in Second Language Proficiency* (pp. 79-94). London: Continuum.

Roehr, K., & Gánem-Gutiérrez, G. A. (2009b). The status of metalinguistic knowledge in instructed adult L2 learning. *Language Awareness, 18*(2), 165-181.

Rogers, V., Meara, P., Aspinall, R., Fallon, L., Goss, T., Keey, E., & Thomas, R. (2016). Testing aptitude. *EUROSLA Yearbook,16*(1), 179-210.

Rubin, J. (1975). What the “good language learner” can teach us. *TESOL Quaterly*, *9*(1), 41-51.

Rutherford, W., & Sharwood Smith, M. (1985). Consciousness-raising and universal grammar. *Applied Linguistics, 6*(3), 274-282.

Sáfár, A., & Kormos, J. (2008). Phonological short-term memory, working memory and foreign language performance in intensive language learning. *Bilingualism: Language* and Cognition, 11(2), 261-271.

Sanz, C. (2005). Adult SLA: The interaction between internal and external factors. In C. Sanz (Ed.), *Mind and context in adult second language acquisition* (pp. 3-20). Georgetown University Press.

Sanz, C., Lin, H., Lado, B., Stafford, C., & Bowden, H. (2014). One size fits all? Learning conditions and working memory capacity in *Ab initio* language development. *Applied Linguistics*, 1-26. Doi:10.1093/applin/amu058.

Sasaki, M. (1996). *Second language proficiency, foreign language aptitude, and intelligence.* New York: Language.

Sawyer, M., & Ranta, L. (2001). Aptitude, individual differences, and instructional design. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 319-53). Cambridge: Cambridge University Press.

Scheffler, P. (2009). Rule difficulty and the usefulness of instruction. *ELT Journal*, *63*(1), 5-12.

Scheffler, P. (2011). Rule difficulty: Teachers' intuitions and learners' performance. *Language Awareness*, *20*(3), 221-237.

Scheffler, P., & Cinciata, M. (2011). Explicit grammar rules and L2 acquisition. *ELT Journal*, *65*(1), 13-23.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, *11*(2), 129-158.

Schmitt, C. J. (2008). *Schaum's outline of Spanish grammar* (5th edition). New York: McGraw-Hill.

Serafini, E. J., & Sanz, C. (2015). Evidence for the decreasing impact of cognitive ability on second language development as proficiency increases. *Studies in Second Language Acquisition*, *38*(2), 1-40.

Service, E. (1992). Phonology, working memory and foreign language learning. *Quaterly Journal of Experimental Psychology*, *45a*, 21-50.

Service E., & Kohonen, V. (1995). Is the relation between phonological memory and foreign language learning accounted for by vocabulary acquisition? *Applied Psycholinguistics, 16*, 155- 172.

Sharwood-Smith, M. (1981). Consciouness-raising and the second language learner. *Applied Linguistics*, *2*, 159-169.

Shiu, L. (2011). *EFL learners' perceptions of grammatical difficulty in relation to second language proficiency, performance, and knowledge* (Unpublished Ph.D. Thesis). University of Ontario.

Skehan, P. (1989). *Individual differences in second language learning*. London: Edward Arnold.

Skehan, P. (1998). *A cognitive approach to language learning.* Oxford: Oxford University Press.

Skehan, P. (2002). Theorising and updating aptitude. In P. Robinson (Ed.), *Individual differences and instructed language learning* (pp. 69-93). Amsterdam: John Benjamins.

Skehan, P. (2012). Language aptitude. In S. Gass, & A. Mackey (Eds.). *The Routledge handbook of second language acquisition.* New York, NY: Routledge Handbooks.

Sorace, A. (1985). Metalinguistic knowledge and language use in acquisition-poor environments. *Applied Linguistics, 6*(3), 239-254.

Spada, N. (2011). Beyond form-focused instruction: Reflections on past, present and future research. *Language Teaching, 44*(2), 225-236.

Spada, N., & Lightbown, P. (2008). Form-focused instruction: Isolated or integrated. *Tesol Quaterly, 42*(2), 181-207.

Spada, N., & Tomita, Y. (2010). Interactions between type of instruction and type of language feature: A meta-analysis. *Language Learning, 60*(2), 263-308.

Spada, N., Li-Ju S., & Tomita, Y. (2015). Validating an elicited imitation task as a measure of implicit knowledge: Comparison with other validation studies. *Language Learning, 65*(3), 723-751.

Suzuki, Y., & DeKeyser, R. (2015). Comparing elicited imitation and word monitoring as measures of implicit knowledge. *Language Learning, 65*(4), 860-895.

Swan, M. (1980). *Practical English Usage*. Oxford: Oxford University Press.

Swan, M. (1995). *Practical English Usage* (2nd Ed.). Oxford: Oxford University Press.

Tellier, A. J. (2015). *Metalinguistic awareness and foreign language learning in primary school: A classroom study with children aged 8 to 9 years* (Unpublished PhD thesis). University of Essex.

Thepseenu, B., & Roehr, K. (2013). University-level learners' beliefs about metalinguistic knowledge. In K. Roehr, & G. A. Gánem-Gutiérrez (Eds.). *The Metalinguistic Dimension in Instructed L2 Learning* (pp. 1-40). London: Continuum.

Trafimov, D., & Marks, M. (2015). Editorial. *Basic and applied social psychology, 37*, 1-2.

Trebits, A., & Kormos, J. (2008). Working memory capacity and narrative task performance. Proceedings from the 33rd International LAUD Symposium, Landau/Pfalz, Germany.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127-145.

Unsworth, N., Heitz, R. P., & Engle, R. W. (2005) Working memory capacity in hot and cold cognition. In R. W. Engle, G. von Hecker Sedek, & D. N. McIntosh (Eds.). *Cognitive limitations in aging and psychopathology* (pp. 19-43). New York: Cambridge University Press.

Van den Noort, M. W. M. L., Bosch, P., & Hugdahl, K. (2006). Foreign language proficiency and working memory capacity. *European Psychologist*, *11*(4), 289-296.

VanPatten, B., & Benati, A. G. (2010). *Key terms in second language acquisition.* London: Continuum.

Waters, G. S., & Caplan, D. (1996). The measurement of verbal WMC and its relation to reading comprehension. *Quaterly Journal of Experimental Psychology*, *Section A: Human Experimental Psychology*, *49A*(1), 51-79.

Wechsler, D. (2003). Weschler intelligence scale for children (4th Ed.) (WISC-IV). *Journal of Psychoeducational Assessment*, 278-295.

Wen, Z. (2012). Working memory and second language learning. *International Journal of Applied Linguistics*, *22*(1), 1-22.

Wen, Z., Mota, M. B., & McNeill, A. (2015). *Working memory in second language acquisition and processing.* Clevedon, UK: Multilingual Matters.

Wesche, M. B. (1981). Language aptitude measures in streaming, matching students with methods, and diagnosis of learning problems. In K. Diller (Ed.), *Individual differences and universals in language aptitude.* Rowley, MA: Newbury House.

Yalçin, S., & Spada, N. (2016). Language aptitude and grammatical difficulty: An EFL classroom based study. *Studies in Second Language Acquisition*, *38*(2), 239-263.

Yalçin, S., Çeçen, S., & Erçetin, G. (2016). The relationship between aptitude and working memory: An instructed SLA context. *Language Awareness*, *25*(1-2), 144-158.

Yilmaz, Y. (2012). Relative effects of explicit and implicit feedback: The role of working memory capacity and language analytic ability. *Applied Linguistics*, *34*(3), 344-368.

Zhang, R. (2015). Measuring university-level L2 learners' implicit and explicit linguistic knowledge. *Studies in Second Language Acquisition*, *37*, 457-486.

Ziętek, A. A., & Roehr, K. (2011). Metalinguistic knowledge and cognitive style in Polish classroom learners of English. *System*, *39*(4), 417-426.

APPENDICES

Appendix A. Background information questionnaire

ID ☐

Name: ____ Date: ________________

Gender: ☐ Male ☐ Female

Bachelor's degree you are currently studying: ____________________________________

Age: _____ years.

How many years have you been learning English? __________ year(s).

Have you learned any other foreign language(s)? ☐ Yes ☐ No

If yes, what foreign language(s) have you learned?

________________________________ How long? ______ years ______ months

________________________________ How long? ______ years ______ months

________________________________ How long? ______ years ______ months

Have you lived in an English speaking country? ☐ Yes ☐ No

If yes, how long? ______ years ______ months Where? ___________________________

1. Do you speak English at home or at work? ☐ Yes ☐ No

   If yes, at home I speak English with ____

   How often? ____

   If yes, at work I speak English with ____

   How often? ____

1. Do you attend the conversation clubs at CAADI? ☐ Yes ☐ No

   If yes, how often? ______________________________

2. Are you going to start the BA in ELT at the Language Department of this university? Yes ☐ No ☐

   If yes, when? ________________________________

3. Are you currently studying in the BA in ELT at the Language Department of this university?

☐ Yes. What semester? ________________________________

☐ No.

Background information questionnaire (Spanish version)

ID ☐

Nombre: ______________________________ Fecha: ______________

Género: ☐ Hombre ☐ Mujer

Carrera que estudias actualmente: ______________________________

Edad: _____ años

¿Por cuánto tiempo has estado estudiando inglés? __________ año(s).

¿Has aprendido algún otro idioma(s) extranjero(s)? ☐ Sí ☐ No

Si es así, ¿Qué otro idioma (s) extranjero(s) has aprendido?

______________________ ¿Por cuánto tiempo? ______ años ______ meses
______________________ ¿Por cuánto tiempo? ______ años ______ meses
______________________ ¿Por cuánto tiempo? ______ años ______ meses
.
Has vivido en un país donde sólo se habla inglés? ☐ Sí ☐ No

Si es así, ¿Por cuánto tiempo? ____ años ____ meses ¿Dónde? ______________

1. ¿Hablas inglés en tu casa o en el trabajo?

   Sí, en mi casa hablo inglés con ______________________________

   ¿Con qué frecuencia? ______________________________

   Sí, en mi trabajo hablo inglés con ______________________________

   ¿Con qué frecuencia? ______________________________

2. ¿Asistes a los clubes de conversación en el CAADI? ☐ Si ☐ No

   Si es así, ¿Con qué frecuencia? ______________________________

3. ¿Vas a comenzar a estudiar la carrera en ELT en el Departamento de Idiomas? Si ☐ No ☐

   Si sí, ¿Cuándo? ______________________

4. ¿Actualmente estás estudiando la carrera en ELT en el Departamento de Idiomas?

¿ ☐ Sí. ¿Qué semestre cursas? ______________________

☐ No

Appendix B. Teacher difficulty judgement questionnaire

ID

<u>Teacher difficulty judgement questionnaire</u>

The questionnaire that you are about to answer is part of an investigation on the learning of various English structures. It is not a test and there are no right or wrong answers. You should give your opinion of each item according to your experience in learning and teaching the language. Your participation is appreciated and of great value to understanding how we teach and learn English.

Instructions:

1. You need to read the statements carefully before you start answering them. (The symbol "*" represents an ungrammatical sentence.]

2. On the right-hand side of the chart, please indicate your opinion on the ***level of learning difficulty*** of each area of English grammar by putting an 'X' in the column that best describes your opinion.

3. If you have any difficulties understanding any of the words, please ask the researcher.

Grammar point	**Pedagogical grammar rule**	**Example sentence(s) (targeted form is in bold)**	**Typical learner error (error is underlined)**	**Learning difficulty**				
				Very easy	Easy	Moderate	Difficult	Very difficult
Simple past tense (*-ed* form)	When a finished action or event in the past is being expressed, the simple past tense is required.	He visit**ed** his brother yesterday.	*He want to buy a new house.					
2^{nd} conditional (*if*-clause)	When an unreal/hypothetical situation is being expressed, the 2^{nd} conditional comprising an *if*-clause with a past tense verb and a main clause with *would* + *infinitive* is used.	If I **had** money, I **would buy** a car.	*If I study more, I would get good grades.					
3^{rd} person –*s* in the simple present tense	When a verb in the 3^{rd} person singular is used in the simple present tense, an –*s* or –*es* is added to the end of the verb of the sentence.	Alex want**s** to go home.	*Maria want to go to the beach.					
Comparative adjectives	When making a comparison, you either add *-er* to a one-syllable adjective or you place *more* in front of an adjective with two or more syllables.	Carlos is tall**er** than his sister. This book is **more** expensive than yours.	*Aguacalientes city is more small than Mexico city. *This garden is more bigger than yours.					
Infinitives and gerunds (as verb complements)	When the main verb of a sentence is, e.g., *decide*, *hope*, or *plan*, and when it is followed by another verb, the *to*-infinitive construction is required for the second verb, but if the main verb is, e.g., *enjoy*, *avoid*, or *deny*, the *ing*-form construction is required for the second verb.	He decided **to write** a story. She enjoys **driving** around the country.	*He wanted travel abroad. *She avoids to talk to strangers.					
Indefinite article	When a countable noun is first mentioned, an indefinite article is required.	They had **a** good class today.	*They bought the new computer.					
Modal verbs + verb	When a modal verb such as *must*, *should*, or *can* is used, it is followed by the infinitive of the main verb.	I should **wait** for my brother.	*I must to do my homework.					

Grammar point	**Pedagogical grammar rule**	**Example sentence(s) (targeted form is in bold)**	**Typical learner error (error is underlined)**	**Learning difficulty**				
				Very easy	Easy	Moderate	Difficult	Very difficult
Many vs. much	When the quantity of something is being referred to, *many* is required for countable nouns and *much* is required for uncountable nouns.	She has **many** activities to do during the school term. They don't have **much** time.	*I drink <u>many</u> water. *I didn't eat <u>much</u> apples today.					
Yes/no questions	When a yes/no question with the auxiliary verb *do* is used, the infinitive of the main verb is required (the auxiliary, not the main verb, is tensed).	Does Maria **like** the new house? Did he **go** to the park?	*<u>Do</u> John swim fast? *Did they <u>went</u> to the museum?					
Plural of nouns	When the plural of a regular noun is being expressed, an –*s* needs to be added to the noun.	It takes a few **minutes** to get to the airport.	The French class starts in five <u>day</u>.					
Since/For	When the specific time of the beginning of an action is expressed, *since* is required, but when the length of time of an action is expressed, *for* is required.	Jane has been in hospital **since** Tuesday. People have used mobile phones **for** many years.	*They have waited <u>for</u> 4 o'clock today. *Children have played games <u>since</u> two hours.					
Direct and indirect objects (Dative alternation)	When an indirect object follows a direct object in a sentence, the preposition *to* is placed in front of the indirect object.	The man gave a letter **to** the boy.	*The woman reported the car accident <u>the police</u>.					
Relative clauses	When a relative clause where the relative pronoun functions as an object is used, a pronoun that makes reference to the subject of the sentence (resumptive pronoun) is not permitted.	**The table** that I saw the other day **is** expensive.	*The dictionary that I used the other day <u>it</u> includes many phrasal verbs.					

Background information

Name: ____

Phone number: ________________________

Email address: ____________________

Gender: ☐ Male ☐ Female

Age: _____ years.

How many years have you been teaching English? __________ year(s).

Appendix C. Learner difficulty judgement questionnaire

ID

Learner difficulty judgement questionnaire

The questionnaire that you are about to answer is part of an investigation on the learning of various English structures. It is not a test and there are no right or wrong answers. You should give your opinion of each item according to your experience in learning the language. Your participation is appreciated and of great value to understanding how we teach and learn English.

Este cuestionario forma parte de una investigación sobre el aprendizaje de varias estructuras en inglés. No es un examen y no hay respuestas correctas o incorrectas. Da tu opinión basándote en tu experiencia en el aprendizaje del idioma. Tu participación es de mucha importancia para el entendimiento sobre el enseñar y aprender el inglés.

Instructions:

1. You do not need to worry about grades in this difficulty judgment questionnaire.
   No habrá calificación alguna en este cuestionario.
2. You need to read the statements carefully before you start answering them. (The symbol "*" represents an ungrammatical sentence).
   Necesitas leer con atención cada enunciado antes de contestar. El símbolo "" representa un enunciado gramaticalmente incorrecto].*
3. On the right-hand side of the chart, please indicate your opinion on the ***level of learning difficulty*** of each area of English grammar by putting an 'X' in the column that best describes your opinion.
   En la parte derecha de la tabla da tu opinión sobre ***el grado de dificultad de aprendizaje*** *de cada área gramatical del inglés tachando el espacio en la columna que mejor describa tu opinión.*
4. Concentrate on your questionnaire.
   Concéntrate en el cuestionario.
5. If you have any difficulties understanding any of the words, please ask the instructor.
   Si no conoces alguna palabra, pregunta al instructor.

Thank you for your participation!

Grammar point	**Pedagogical grammar rule**	**Example sentence(s) (targeted form is in bold)**	**Typical learner error (error is underlined)**	**Learning difficulty**				
				Very easy	Easy	Moderate	Difficult	Very difficult
Simple past tense (*-ed* form)	When a finished action or event in the past is being expressed, the simple past tense is required.	He visit**ed** his brother yesterday.	*He want to buy a new house.					
2nd conditional (*if*-clause)	When an unreal/hypothetical situation is being expressed, the 2nd conditional comprising an *if*-clause with a past tense verb and a main clause with *would + infinitive* is used.	If I **had** money, I **would buy** a car.	*If I study more, I would get good grades.					
3rd person *–s* in the simple present tense	When a verb in the 3rd person singular is used in the simple present tense, an *–s* or *–es* is added to the end of the verb of the sentence.	Alex want**s** to go home.	*Maria want to go to the beach.					
Comparative adjectives	When making a comparison, you either add *-er* to a one-syllable adjective or you place *more* in front of an adjective with two or more syllables.	Carlos is tall**er** than his sister. This book is **more** expensive than yours.	*Aguacalientes city is more small than Mexico city. *This garden is more bigger than yours.					
Infinitives and gerunds (as verb complements)	When the main verb of a sentence is, e.g., *decide*, *hope*, or *plan*, and when it is followed by another verb, the *to*-infinitive construction is required for the second verb, but if the main verb is, e.g., *enjoy*, *avoid*, or *deny*, the *ing*-form construction is required for the second verb.	He decided **to write** a story. She enjoys **driving** around the country.	*He wanted travel abroad. *She avoids to talk to strangers.					
Indefinite article	When a countable noun is first mentioned, an indefinite article is required.	They had **a** good class today.	*They bought the new computer.					
Modal verbs + verb	When a modal verb such as *must*, *should*, or *can* is used, it is followed by the infinitive of the main verb.	I should **wait** for my brother.	*I must to do my homework.					

Grammar point	Pedagogical grammar rule	Example sentence(s) (targeted form is in bold)	Typical learner error (error is underlined)	Learning difficulty				
				Very easy	Easy	Moderate	Difficult	Very difficult
Many vs. much	When the quantity of something is being referred to, *many* is required for countable nouns and *much* is required for uncountable nouns.	She has **many** activities to do during the school term. They don't have **much** time.	*I drink <u>many</u> water. *I didn't eat <u>much</u> apples today.					
Yes/no questions	When a yes/no question with the auxiliary verb *do* is used, the infinitive of the main verb is required (the auxiliary, not the main verb, is tensed).	Does Maria **like** the new house? Did he **go** to the park?	*<u>Do</u> John swim fast? *Did they <u>went</u> to the museum?					
Plural of nouns	When the plural of a regular noun is being expressed, an –*s* needs to be added to the noun.	It takes a few **minutes** to get to the airport.	The French class starts in five <u>day</u>.					
Since/For	When the specific time of the beginning of an action is expressed, *since* is required, but when the length of time of an action is expressed, *for* is required.	Jane has been in hospital **since** Tuesday. People have used mobile phones **for** many years.	*They have waited <u>for</u> 4 o'clock today. *Children have played games <u>since</u> two hours.					
Direct and indirect objects (Dative alternation)	When an indirect object follows a direct object in a sentence, the preposition *to* is placed in front of the indirect object.	The man gave a letter **to** the boy.	*The woman reported the car accident <u>the police</u>.					
Relative clauses	When a relative clause where the relative pronoun functions as an object is used, a pronoun that makes reference to the subject of the sentence (resumptive pronoun) is not permitted.	**The table** that I saw the other day **is** expensive.	*The dictionary that I used the other day <u>it</u> includes many phrasal verbs.					

Appendix D. Elicited imitation test

ID

This is a beliefs questionnaire. I am going to ask you your opinion about a range of topics.
Este es un cuestionario de opinión. Te pido tu opinión en varios temas.

You will hear a statement. Decide whether the statement is true/not true for you or whether you are not sure. Circle the option to indicate whether you think the statement is true, not true or whether you are not sure. Then repeat the statement in correct English.
Escucharás un enunciado. Luego decide si para ti el enunciado es verdadero/no es verdadero o si no estás seguro(a). Encierra en un círculo la opción que creas relevante. Después repite el enunciado en inglés correcto.

Training section

Statement A
You will hear: “Life is very difficult for old people.” True Not sure Not sure
Now circle one of the options.
[PAUSE]
Now repeat the statement.
[PAUSE]
What you should have said is Life is very difficult for old people.

Statement B
You will hear: “English spoken in many different countries” True Not true Not sure
Now circle one of the options.
[PAUSE]
Now repeat the statement.
[PAUSE]
What you should have said is English is spoken in many different countries.

Now here are some more statements for you to practice with. Remember you are to decide whether each statement is true/not true for you or whether you are not sure. Then you are to repeat the statement in correct English.

Aquí hay más enunciados para que sigas practicando. Recuerda que tienes que decidir si para ti el enunciado es verdadero/no es verdadero o si no estás seguro(a). Encierra en un círculo la opción que creas que es verdadera/no es verdadera o si no estás seguro(a). Después repite el enunciado en inglés correcto.

Statement C

You will hear: "Young people watch television and don't read books."

True Not true Not sure

Now circle one of the options.

[PAUSE]

Now repeat the statement.

[PAUSE]

What you should have said is Young people watch television and don't read books.

Statement D

You will hear: "A good doctor always listens what patients say." True Not true Not sure

Now circle one of the options.

[PAUSE]

Now repeat the statement.

[PAUSE]

What you should have said is A good doctor always listens to what patients say.

Statement E

You will hear: "If you like good food, you should eat always at McDonalds."

True Not true Not sure

Now circle one of the options.

[PAUSE]

Now repeat the statement.

[PAUSE]

What you should have said is If you like good food you should always eat at McDonalds.

Statement F

You will hear: “The invention of the aeroplane has changed the world.”

True Not true Not sure

Now circle one of the options.

[PAUSE]

Now repeat the statement.

[PAUSE]

What you should have said is The invention of the aeroplane has changed the world.

The training is now finished. Please turn over your page and start the questionnaire.

Aquí se termina el entrenamiento. Favor de dar vuelta a la página para comenzar con el cuestionario.

Beliefs questionnaire

1) “Mexico is **greener** and **more beautiful** than other countries.” True Not true Not sure
[PAUSE]
Now repeat the statement.

2) “Mexican people want **to keep** their country clean and green.” True Not true Not sure
[PAUSE]
Now repeat the statement.

*3) “People should report stolen money the police.” True Not true Not sure
[PAUSE]
Now repeat the statement.

*4) “Everyone loves comic books and read them.” True Not true Not sure
[PAUSE]
Now repeat the statement.

5) “The film that everyone likes **is** Star Wars.” True Not true Not sure
[PAUSE]
Now repeat the statement.

6) “People can **win** a lot of money in a casino.” True Not true Not sure
[PAUSE]
Now repeat the statement.

7) “People should report a car accident **to** the police.” True Not true Not sure
[PAUSE]
Now repeat the statement.

8) “People have used computers since many years.” True Not true Not sure
[PAUSE]
Now repeat the statement.

*9) “The software that Bill Gates invented it changed the world.” True Not true Not sure
[PAUSE]
Now repeat the statement.

10) “A bad teacher **makes** lessons interesting and **cares** about students.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*11) “Not everyone can to learn a second language.” True Not true Not sure

[PAUSE]

Now repeat the statement.

12) “To speak English well you must study **for** one year.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*13) “It is more harder to learn Japanese than to learn English.” True Not true Not sure

[PAUSE]

Now repeat the statement.

14) “Mijares **loved** Lucero but he **divorced** her.” True Not true Not sure

[PAUSE]

Now repeat the statement.

15) “If John Lennon **were** alive today, many people **would be** happy.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*16) “Young boys like fast car.” True Not true Not sure

[PAUSE]

Now repeat the statement.

17) “If Russia **had** more power, the United States **would be** worried.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*18) “When man invented the motor car, life change for everyone.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

19) “People need **many** skills to learn English.” True Not true Not sure

[PAUSE]

Now repeat the statement.

20) “Parents have **a** responsibility to care for their children.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*21) “Rich people have two or three house.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*22) “Every child needs good father.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*23) “It is a silly question to ask, ‘Do a woman need to marry?’” True Not true Not sure

[PAUSE]

Now repeat the statement.

*24) “People in love usually want get married as soon as possible.” True Not true Not sure

[PAUSE]

Now repeat the statement.

25) “It is difficult to ask, ‘Do you really **love** me?’” True Not true Not sure

[PAUSE]

Now repeat the statement.

26) “**Many** people study at university level today.” True Not true Not sure

[PAUSE]

Now repeat the statement.

27) “Einstein **failed** Math when he was a student.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*28) “A good student never study before an exam.” True Not true Not sure

[PAUSE]

Now repeat the statement.

29) “Children don’t enjoy **going** to the beach in the summer.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*30) “Poor people need many money to travel” True Not true Not sure

[PAUSE]

Now repeat the statement.

31) “University **students** don’t have many **books**.” True Not true Not sure

[PAUSE]

Now repeat the statement.

32) “People have played sports **for** many years.” True Not true Not sure

[PAUSE]

Now repeat the statement.

33) “Students should give feedback **to** teachers.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*34) “The city that many people want to visit it is Paris.” True Not true Not sure

[PAUSE]

Now repeat the statement.

35) “Zacatecas is a **nicer** place to visit than Leon.” True Not true Not sure

[PAUSE]

Now repeat the statement.

36) “If politicians **were** interested in people’s problems, life **would be** different.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

37) “Two Mexican scientists **discovered** a new planet last year.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*38) “It is a silly question to ask ‘Does a student needs to study?’” True Not true Not sure

[PAUSE]

Now repeat the statement.

39) “English teachers don’t have **much** work.” True Not true Not sure

[PAUSE]

Now repeat the statement.

40) “Teachers must **prepare** their classes before they give a lesson.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*41) “If languages were easy to learn, people will study on their own.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*42) “Hundreds of people visit Cancun last summer.” True Not true Not sure

[PAUSE]

Now repeat the statement.

43) “Eros Ramazzotti **sings** pop music.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*44) “Students can’t avoid to ask about the exams.” True Not true Not sure

[PAUSE]

Now repeat the statement.

45) “It is an interesting question to ask ‘Does a teacher **want** a better job?’”

True Not true Not sure

[PAUSE]

Now repeat the statement.

46) “It is not a good idea for **teachers** to punish **students**.” True Not true Not sure

[PAUSE]

Now repeat the statement.

47) “President Peña Nieto has been in the presidency **since** 1994.” True Not true Not sure

[PAUSE]

Now repeat the statement.

48) “The language that most people speak **is** French.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*49) “You should give a present your mother on her birthday.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*50) “It is an interesting question to ask ‘Do you likes living in Aguascalientes?’ ”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*51) “People must to reserve a hotel before going to the beach.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*52) “If Europe were closer to Latin America, more people will visit it.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

53) “Basketball is **more popular** than soccer around the world.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*54) “Most young people enjoy to listen to music.” True Not true Not sure

[PAUSE]

Now repeat the statement.

55) “Soccer players must **warm up** before the game starts.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*56) “Much students know Mexican history.” True Not true Not sure

[PAUSE]

Now repeat the statement.

57) “It is a silly question to ask, ‘Does a child **need** education?’ ” True Not true Not sure

[PAUSE]

Now repeat the statement.

*58) “A medicine student needs to study for five year to become a doctor.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*59) “Chicharito has played with Manchester United for 2006.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*60) “The government should give financial aid poor people.” True Not true Not sure

[PAUSE]

Now repeat the statement.

61) “The two sports that most people watch **are** soccer and baseball.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*62) “Thousands of people attend the World Cup in Brazil this year.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*63) “If soccer is not as popular as it is today, people would watch other sports on TV.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

64) “Chicharito **plays** soccer with Chivas Rayadas.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*65) “Vicente Fernandez is more rich than Emmanuel.” True Not true Not sure

[PAUSE]

Now repeat the statement.

66) “Most famous singers started **singing** at a young age.” True Not true Not sure

[PAUSE]

Now repeat the statement.

67) “Every city has **an** interesting museum to visit.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*68) “Students should to study hard if they want to get a good grade.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*69) “High school students don’t need many time to study.” True Not true Not sure

[PAUSE]

Now repeat the statement.

70) “You need to go to the library to buy the good book.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*71) “The two political parties that the United States has they are the Republicans and the Democrats.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

72) “Parents should give a good education **to** their children.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*73) “President Peña Nieto speak in public very often.” True Not true Not sure

[PAUSE]

Now repeat the statement.

74) “Music is **a** hard subject to study.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*75) “Learning English is more easier than learning French.” True Not true Not sure

[PAUSE]

Now repeat the statement.

76) “Bus **drivers** don’t work more than 8 **hours** every day.” True Not true Not sure

[PAUSE]

Now repeat the statement.

*77) “Vicente Fernandez has been the most popular ‘rancheras’ singer since three decades.”

True Not true Not sure

[PAUSE]

Now repeat the statement.

*78) “Paris is the good place to visit during vacations.” True Not true Not sure

[PAUSE]

Now repeat the statement.

Appendix E. Oral narrative test

ID

Every morning Mr. García **gets up** at 6:30 am, **walks** to **a** store and **buys a** newspaper. He **returns** to his house and **reads** the newspaper. Then, if he **feels** like it, he **goes** to work. But often he **stays** at home, **drinks a** cup of coffee and **sits** in the sun. Some **days** are **hotter** than **others**. His wife **says** he **should go** to work. But he always **smiles** and **says**, "I **want to take** life easy. I **want to enjoy** myself. Every day I **enjoy reading** the newspaper. I also **like sitting** in the sun".

Yesterday something **happened** to Mr. García. This is what **happened**.

Mr. García found **a** wallet. It **contained** some **bills** and **coins**, some **credit cards** and two **lottery tickets**. One ticket was **older and more valuable than** the other one. Mr. García **checked** the lottery ticket's number in the newspaper. He **couldn't believe** it. He had the winning ticket. It was worth 50 million **dollars**.

Mr. García didn't know what to do. He **showed** the ticket **to his wife**. She told him he **should** return the ticket **to the owner**?". He took the bus to the address in the wallet. He **knocked** on the door. **A** woman **younger than** him **opened** the door.

"**Can** I see Mr. Romo?," **asked** Mr. García.

"Just **a** minute. He is my brother", said the woman. **An older** man **than** the woman came to the door.

Mr. Romo came to the door. Mr. García **showed** the newspaper **to the man**. He also gave the ticket **to the man**. "This ticket **belongs** to you", he said.

Mr. Romo **couldn't believe** that he had won the lottery. He thought it was **a** dream!

"I **want to thank** you for being **more honest than** most people I know", he said. "I **want to give** some money **to you**. **Would** you **mind taking** five million **dollars**? Is that enough?"

Mr. García **accepted** the five million **dollars**. His life **changed**. He no longer **needed to work**. In fact he and his wife **enjoyed spending** the money on buying new **houses** and **visiting** many **places** around the world.

Appendix F. Metalinguistic knowledge test (Part 1)

ID

Instructions:

Each of the following 18 items contains an instance of unacceptable use of English. Please correct the highlighted mistake and then describe and explain why the underlined part represents a mistake. You can give your description and explanation in English or Spanish. Please **provide as much detail as possible** when describing and explaining your correction of each sentence.

En cada una de las siguientes oraciones hay una parte que no es correcta. Por favor corrige el error que está resaltado en la oración y después describe y explica el porqué del error. Puedes dar tu descripción y explicación en inglés o en español. Por favor trata de describir y explicar el porqué del error **con el mayor detalle posible**.

Examples:

a) If I <u>have had</u> enough money last year, I would have bought a house.

Correction: had had

Explanation: When a past situation that did not happen is being referred to, past perfect tense is required.

b) The crime <u>solved</u> last week.

Correction: was solved

Explanation: When the performer of an action is unknown, passive voice is required.

You will find that some items are harder than others, although they are not presented in order of difficulty. Please **provide as much detail as possible** when describing and explaining your correction of each sentence.

Algunas de las oraciones son más difíciles que otras pero no están ordenadas en orden de dificultad. Por favor trata de describir y explicar el porqué del error **con el mayor detalle posible**.

1. When he finished his homework, he <u>watch</u> a movie.

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………

…………………………………………………….………………………………………………….

2. If I <u>know</u> the answer, I would tell you.

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………

……………………………………………………………………………………………………..

3. Sara <u>cook</u> every day.

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………

……………………………………………………………………………………………………..

4. This car is <u>more cheaper</u> than mine.

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………

……………………………………………………………………………………………………..

5. Mike is <u>more tall</u> than Joe.

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………

……………………………………………………………………………………………………..

6. The boys want <u>buy</u> a new car.

Correction: ………………………

Explanation:

…………………………………………………………..………………………………………

………………………………………………………………………………………………..

7. They finished <u>to build</u> the house.

Correction: ………………………

Explanation:

…………………………………………………………..………………………………………

………………………………………………………………………………………………..

8. She bought <u>the</u> new house.

Correction: ………………………

Explanation:

…………………………………………………………..………………………………………

…..……………………………………………………………………………………………

9. I must <u>to go</u> to work.

Correction: ………………………

Explanation:

…………………………………………………………..………………………………………

………………………………………………………………………………………………..

10. I have <u>many</u> money.

Correction: ………………………

Explanation:

…………………………………………………………..………………………………………

………………………………………………………………………………………………..

11. I didn't see <u>much</u> people at school today.

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………
……………………………………………………………………………………………..

12. <u>Do</u> Pedro work late?

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………
……………………………………………………………………………………………..

13. Did they <u>took</u> the book?

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………
……………………………………………………………………………………………..

14. Joe sold his two <u>car</u> to a friend.

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………
……………………………………………………………………………………………..

15. I have been here <u>for</u> 9 o'clock.

Correction: ………………………

Explanation:

…………………………………………………………..…………………………………………
……………………………………………………………………………………………..

16. Teachers have used computers <u>since</u> two decades.

Correction: ………………………

Explanation:

………………………………………………………..………………………………………
…………………………………………………………………………………………………..

17. The postman gave the letter <u>the woman</u>.

Correction: ………………………

Explanation:

………………………………………………………..………………………………………
…………………………………………………………………………………………………..

18.The car that my father bought <u>it</u> is new.

Correction: ………………………

Explanation:

………………………………………………………..………………………………………
…………………………………………………………………………………………………..

Metalinguistic knowledge test (Part 2)

ID

Instructions:

Write a sentence for each of the following 17 rules. You can write your sentences in positive, negative or interrogative form. You should write your sentences in correct English. Write your sentence according to the rule provided. In the examples below you will see two sentences: the first sentence covers all aspects of the rule and the second sentence does not.

Escribe una oración para cada una de las siguientes 17 reglas gramaticales. Las oraciones las puedes escribir en forma positiva, negativa o interrogativa. Escribe tus oraciones en inglés correcto. Escribe tus oraciones de acuerdo a lo que dice la regla gramatical. En los ejemplos que siguen la primera oración obedece a la regla y el segundo ejemplo es insuficiente.

Examples:

Grammar point: **Passive voice**

a) **Rule:** When the agent is mentioned in a passive voice construction, the word *by* is put in front of the agent.

Cuando el agente es mencionado en una oración en la voz pasiva, la palabra "by" se pone en frente del agente.

Sentence (covering all aspects of the rule): <u>The classrooms were painted by the students.</u>

Sentence (**not** covering all aspects of the rule): <u>The classrooms were painted.</u>

The second sentence does not cover all aspects of the rule because it is missing the word *by* and the agent.

La segunda oración no cubre todos los aspectos de la regla gramatical porque le falta la palabra "by" y el agente.

Grammar point: **Present perfect tense**

b) **Rule:** When an action happened a short time ago, the present perfect is required.

Cuando una acción acaba de suceder, se requiere el presente perfecto.

Sentence (covering all aspects of the grammar rule): I have just finished studying for my English exam.

Sentence (**not** covering all aspects of the rule): I have finished studying for my English exam.

The second sentence does not cover all aspects of the rule because it is missing the word *just* which indicates that the action happened a short time ago.

La segunda oración no cubre todos los aspectos de la regla gramatical porque le falta la palabra "just" la cual indica que la acción sucedió hace poco tiempo.

You will find that some rules are harder than others, although they are not presented in order of difficulty. Please try your best to write a sentence covering **all** aspects of the grammar rule. Encontrarás que algunas reglas gramaticales son más difíciles que otras pero no están ordenadas en orden de dificultad. Trata de escribir las oraciones tratando de cubrir **todos** los aspectos de la regla gramatical.

Grammar point: **Simple past tense (*-ed* form)**

1. **Rule:** When a finished action or event in the past is being expressed, the simple past tense is required. (Please use a regular verb).
   Cuando se expresa un evento u acción terminada en el pasado, se utiliza el pasado simple. (Favor de usar un verbo regular en la oración).

Sentence: ………………………………………………………………………………………

………………………………………………………………………………………

Grammar point: **2nd conditional (*if*-clause)**

2. **Rule:** When an unreal/hypothetical situation is being expressed, the 2nd conditional comprising an *if*-clause with a past tense verb and a main clause with *would + infinitive* is used. *Cuando se expresa una situación irreal/hipotética, se usa el Segundo Condicional que consta de la clásula "if-clause" con el verbo en pasado y la cláusula principal con "would + infinitive".*

Sentence: ………………………………………………………………………………………

………………………………………………………………………………………

Grammar point: **3rd person –*s* in the simple present tense**

3. **Rule:** When a verb in the 3rd person singular is used in the simple present tense, an *–s* or *–es* is added to the end of the main verb of the sentence.
   Cuando se usa un verbo con la 3ra persona del singular en tiempo presente simple, se le añade "-s" o "-es".

Sentence: ………………………………………………………………………………………

………………………………………………………………………………………

Grammar point: **Comparative adjectives**

4. **Rule:** When a comparative is formed for a one-syllable adjective, *-er* is added.
   Cuando se forma una forma comparativa para un adjetivo de una sílaba, se añade "-er".

Sentence: ………………………………………………………………………………………

………………………………………………………………………………………

5. **Rule:** When a comparative is formed for an adjective with two or more syllables, *more* is placed in front.
   Cuando se forma una forma comparativa para un adjetivo con dos o más sílabas, se coloca "more" antes del adjetivo.

Sentence: ………………………………………………………………………………………

………………………………………………………………………………………

Grammar point: **Infinitives and gerunds (as verb complements)**

6. Rule: When the main verb of a sentence is, e.g., *decide*, *hope*, or *plan*, and when it is followed by another verb, the *to*-infinitive construction is required for the second verb.

Cuando el verbo principal de una oración como "decide", "hope" o "plan" le sigue otro verbo, el segundo verbo tiene que estar en infinitivo.

Sentence: ………………………………………………………………………………………

………………………………………………………………………………………

7. Rule: When the main verb of a sentence is, e.g., *enjoy*, *avoid*, or *deny*, and when it is followed by another verb, the *ing*-form construction is required for the second verb.

Cuando el verbo principal de una oración como "enjoy", "avoid", or "deny" le sigue otro verbo, el segundo verbo tiene que estar en gerundio.

Sentence: ………………………………………………………………………………………

………………………………………………………………………………………

Grammar point: **Indefinite article**

8. **Rule:** When a countable noun is first mentioned, an indefinite article is required.

Cuando se menciona un sustantivo por primera vez, se requiere un artículo indefinido.

Sentence: ………………………………………………………………………………………

………………………………………………………………………………………

Grammar point: **Modal verbs + verb (in simple form)**

9. Rule: When a modal verb such as *must*, *should*, or *can* is used, it is followed by the infinitive of the main verb.

Cuando se usa un verbo modal como "must", "should", o "can", el verbo que le acompaña se pone en forma infinitiva.

Sentence: ………………………………………………………………………………………

………………………………………………………………………………………

Grammar point: **Many vs. much**

10. Rule: When the quantity of something is being referred to, *many* is required for countable nouns.

Cuando se hace referencia a la cantidad de algo, se requiere "many" para sustantivos contables.

Sentence: ………………………………………………………………………………………
………………………………………………………………………………………

11. Rule: When the quantity of something is being referred to, *much* is required for uncountable nouns.

Cuando se hace referencia a la cantidad de algo, se requiere "much" para sustantivos no contables.

Sentence: ………………………………………………………………………………………
………………………………………………………………………………………

Grammar point: Yes/no questions

12. Rule: When a yes/no question with the auxiliary verb *do* is used, the infinitive of the main verb is required (the auxiliary, not the main verb, is tensed)

Cuando se usa el verbo auxiliar "do" en una pregunta de "yes/no", se requiere de la forma infinitiva del verbo principal (el auxiliar, no el verbo principal, es el que indica el tiempo de la oración).

Sentence: ……………………………………………………………………………………………..
……………………………………………………………………………………………..

Grammar point: **Plural of nouns**

13. **Rule:** When the plural of a regular noun is being expressed, an *–s* needs to be added to the noun.

Cuando se expresa un sustantivo regular en plural, se añade una "s" al sustantivo.

Sentence: ………………………………………………………………………………………
………………………………………………………………………………………

Grammar point: **Since/For**

14. **Rule:** When the specific time of the beginning of an action is expressed, *since* is required.

Cuando se expresa el tiempo específico del inicio de una acción, se usa "since".

Sentence: ……………………………………………………………………………………

……………………………………………………………………………………

15. **Rule:** When the length of time of an action is expressed, *for* is required.

Cuando se expresa un periodo de tiempo, se usa "for".

Sentence: ……………………………………………………………………………………

……………………………………………………………………………………

Grammar point: **Direct and indirect objects (Dative alternation)**

16. Rule: When an indirect object follows a direct object in a sentence, the preposition *to* is placed in front of the indirect object.

Cuando en una oración un objeto indirecto acompaña a un objeto directo, se coloca la preposición "to" al frente del objeto directo.

Sentence: ……………………………………………………………………………………

……………………………………………………………………………………

Grammar point: **Relative clauses**

17. Rule: When a relative clause where the relative pronoun functions as an object is used, a pronoun that makes reference to the subject of the sentence (resumptive pronoun) is not permitted.

Cuando se usa una cláusula relativa donde el pronombre relativo funciona como un objeto, no se permite repetir el pronombre que hace referencia al sujeto de la oración.

Sentence: ……………………………………………………………………………………

……………………………………………………………………………………

Appendix G. Backward digit span test

ID

		Trial 1	Trial 2
3	5 - 8 – 2		
	6 - 9 – 4		
	1 - 4 – 8		
	2 - 7 – 6		
4	6 - 4 - 3 – 9		
	7 - 2 - 8 – 6		
	9 - 6 - 2 – 5		
	7 - 4 - 9 – 1		
5	4 - 2 - 7 - 3 – 1		
	7 - 5 - 8 - 3 – 6		
	6 - 4 - 7 - 8 – 1		
	9 - 6 - 2 - 7 – 4		
6	6 - 1 - 9 - 4 - 7 – 3		
	3 - 9 - 2 - 4 - 8 – 7		
	7 - 1 - 8 - 4 - 9 – 5		
	1 - 5 - 7 - 4 - 2 – 9		
7	5 - 9 - 1 - 7 - 4 - 2 – 8		
	4 - 1 - 7 - 9 - 3 - 8 – 6		
	6 - 5 - 1 - 7 - 4 - 9 – 2		
	1 - 4 - 7 - 5 - 3 - 8 – 6		
8	5 - 8 - 1 - 9 - 2 - 6 - 4 – 7		
	3 - 7 - 2 - 9 - 5 - 1 - 8 – 4		
	5 - 9 - 1 - 6 - 8 - 3 - 4 - 2		
	3 - 2 - 5 - 7 - 4 - 9 - 1 – 8		
9	2 - 7 - 5 - 8 - 6 - 2 - 9 - 1 - 4		
	7 - 1 - 3 - 9 - 4 - 2 - 5 - 6 - 8		
	8 - 1 - 3 - 9 - 6 - 2 - 5 - 7 - 4		
	2 - 9 - 5 - 1 - 7 - 3 - 4 - 6 – 8		

Appendix H. Consent forms.

Consent Form for Student Participants

Department of Language and Linguistics, University of Essex
Supervisor: Dr Karen Roehr-Brackin
Researcher: Luis Humberto Rodríguez Silva

What is the project about?

The project investigates Spanish-speaking students' knowledge of selected English structures and how such knowledge relates to language learning aptitude and working memory. It also looks at students' and teachers' beliefs about the difficulty of the English structures under study.

What does participating involve?

It involves completing one beliefs questionnaire, a difficulty judgement questionnaire, two language tests, one aptitude test, and a working memory test.

Participation will not take more than three hours maximum. The instruments will be administered in different sessions.

Please tick the appropriate boxes	Yes	No
Taking Part		
I have read and understood the project information given above.	☐	☐
I have been given the opportunity to ask questions about the project.	☐	☐
I agree to take part in the project. Taking part in the project will involve completing two questionnaires, two language tests, one aptitude test, and a working memory test.	☐	☐
I agree to take part in the questionnaire and the language test which will be audio-recorded.	☐	☐
I understand that my taking part is voluntary; I can withdraw from the study at any time and I do not have to give any reasons for why I no longer want to take part.	☐	☐
Use of the information I provide for this project only		
I understand my personal details such as name, email address and phone number will not be revealed to people outside the project.	☐	☐
I understand that my words may be quoted in the dissertation which will report on this project.	☐	☐

________________________	________________________	________
Name of participant	Signature	Date
________________________	________________________	________
Researcher	Signature	Date

Contact details for further information:

Researcher: LUIS HUMBERTO RODRIGUEZ SILVA; E-mail: lhrodr@essex.ac.uk; Mobile phone:07415440575

Supervisor: DR. KAREN ROEHR-BRACKIN; E-mail: kroehr@essex.ac.uk

Consent Form for Teacher Participants

Department of Language and Linguistics, University of Essex
Supervisor: Dr Karen Roehr
Researcher: Luis Humberto Rodríguez Silva

What is the project about?

The project investigates Spanish-speaking students' knowledge of selected English structures and how such knowledge relates to language learning aptitude and working memory. It also looks at students' and teachers' beliefs about the difficulty of the English structures under study.

What does participating involve?

It involves completing a difficulty judgement questionnaire.

Participation will not take more than 20 minutes.

Please tick the appropriate boxes	**Yes**	**No**
Taking Part		
I have read and understood the project information given above.	☐	☐
I have been given the opportunity to ask questions about the project.	☐	☐
I agree to take part in the project. Taking part in the project will involve completing a questionnaire.	☐	☐
I understand that my taking part is voluntary; I can withdraw from the study at any time and I do not have to give any reasons for why I no longer want to take part.	☐	☐
Use of the information I provide for this project only		
I understand my personal details such as name, email address and phone number will not be revealed to people outside the project.	☐	☐
I understand that my words may be quoted in the dissertation which will report on this project.	☐	☐

________________________	________________________	________
Name of participant	Signature	Date
________________________	________________________	________
Researcher	Signature	Date

Contact details for further information:

Researcher: LUIS HUMBERTO RODRIGUEZ SILVA; E-mail: lhrodr@essex.ac.uk; Mobile phone:07415440575

Supervisor: DR. KAREN ROEHR-BRACKIN; E-mail: kroehr@essex.ac.uk

Appendix I. Scoring scheme for the MLK test

Metalinguistic knowledge test (part 1)

Correction and description/explanation

Adequate correction = 1 point
Adequate description (what form?) = 1 point
Adequate explanation (why this form?) = 1 point

Metalinguistic knowledge test (part 2)

Sentence production

Adequate description (what form?) = 1 point
Adequate context (adding to meaning of sentence) = 1 point

Scoring scheme: Correction, description, explanation

A description should answer the question 'What form?'. It is deemed adequate if it is not incorrect and if there is at least some evidence of meaningful generalization beyond the instance provided in the item that is being described. Therefore, an adequate description will usually include appropriate use of at least some metalinguistic terminology.

An explanation should answer the question 'Why this form?'. It is deemed adequate if it is not incorrect and if there is at least some evidence of meaningful generalization beyond the instance provided in the item that is being explained. Therefore, an adequate explanation will usually include appropriate use of at least some metalinguistic terminology.

As far as possible, the answer key provides targeted answers in the default format of a prescriptive pedagogical grammar rule: "When X occurs / function X is being expressed (= explanation), form Y needs to be used (= description)". It is not necessary for informants to fully achieve the descriptive and explanatory detail of the answer key to be awarded a point (see above).

Examples:

Q1:Target: Simple past tense

1. When he finished his homework, he <u>watch</u> a movie.
Correction: watched

Explanation: When a finished action or event in the past is being expressed, the simple past tense is required.

Q1: S26

Explanation: "It is a past action."

Description = 0
Explanation = 1

Q5: Target: Comparative adjectives

5. Mike is more tall than Joe.

Correction: taller
Explanation: When a comparative is formed for a one-syllable adjective, “-er” is added.

Q5: S27

Explanation: “Add ‘er’ in words of 2 syllables”.

Description = 1
Explanation = 0

Q6: Target: Verb complements

6. The boys want buy a new car.

Correction: to buy
Explanation: When the main verb of a sentence such as “decide”, “hope” or “plan” is followed by another verb, the *to*-infinitive construction is required for the second verb.

Q6: S29

Explanation: “After the verb *want* we need to”
Description = 1
Explanation = 1

Rationale / analysis of scores:

- Correction score to be treated separately because corrections can be based on implicit knowledge; task is included because pilot studies showed that correction is a natural step preceding description/explanation of an error.

- Description/explanation are treated separately because answers are likely to be complex with qualitative differences between them: the more we can quantify qualitative differences, the better.

(Section I: Correction)

0	The participant failed to correct the error.
1	The participant corrected the error.

(Section I: Description/explanation)

DESCRIPTION (what form?)	0	• The participant was unable to describe the correction of the error. -The participant provides the incorrect form of the structure needed to correct the error. -The participant's description of the correction of the error is imprecise or incomplete. -The participant translated the stimulus sentence instead of describing or explaining the correction of the error. -The participant provided two options in the explanation or description of the error.
	1	• The participant's description of the correction of the error is correct. -The participant provides the correct form of the structure needed to correct the error. -The participant provides an acceptable rule with some metalanguage. -The participant provides the required rule with appropriate metalanguage.
EXPLANATION	0	• The participant was unable to explain the correction of the error. -The participant inaccurately explains why the form of the structure to correct the error is needed. -The participant's explanation of the correction of the error is imprecise or incomplete.
	1	• The participant's explanation of the correction of the error is correct. -The participant accurately explains why the form of the structure to correct the error is needed. -The participant provides some metalanguage in the explanation. -The participant provides appropriate metalanguage in the explanation.

(Section II: Sentence production)

0	The participant failed to provide an example, or the example they provided was incorrect or was in the wrong context (where necessary)
1	The participant used the target rule in a correct example, but without a clear context (where necessary).
2	The participant gave a completely correct example with an appropriate context (where necessary).

Appendix J. Scoring of implicit and explicit measures

	Implicit measures				Explicit measures			
	Elicited imitation test		Oral narrative		Metalinguistic knowledge			
Grammar point	**Elicited statements (*Ungrammatical sentences)**	**Max poss.**	**Elicited statements**	**Max poss.**	**Statements**	**Max Possible**		
						C	D/E	W/S
Simple past tense (-*ed* form)	14. Mijares **loved** Lucero but he **divorced** her. 27. Einstein **failed** Math when he was a student. 37. Two Mexican scientists **discovered** a new planet last year. *18. When man invented the motor car, life change for everyone. *42. Hundreds of people visit Cancun last summer. *62. Thousands of people attend the World Cup in Brazil this year.	2 1 1 2 1 1 Subtotal = 8	1. Yesterday something **happened** to Mr. Garcia. 2. This is what **happened**. 3. It **contained** 200 pesos. 4. Mr. Garcia **checked** the numbers of the lottery tickets. 5. He **showed** the ticket to his wife. 6. He **knocked** on the door. 7. A woman younger than him **opened** the door. 8. "Can I see Mr. Romo?," **asked** Mr. Garcia. 9. Mr. Garcia **showed** the newspaper to the man. 10. Mr. Garcia **accepted** the five million dollars. 11. His life **changed**. 12. He no longer **needed** to work. 13. In fact he and his wife **enjoyed** spending the money…	1 1 1 1 1 1 1 1 1 1 1 1 1 Subtotal =13	(Part 1) 1.When he finished his homework, he watch a movie. (Part 2) Grammar point: **Simple past tense (-*ed* form)** 1. **Rule:** (It requires context)	1	2	1 Subtotal = 4

2nd conditional (*if*-clause)	15. If John Lennon **were** alive today, many people **would be** happy. 17. If Russia **had** more power, the United States **would be** worried. 36. If politicians **were** interested in people's problems, life **would be** different. *41. If languages were easy to learn, people <u>will</u> study on their own. *52. If Europe were closer to Latin America, more people <u>will</u> visit it. *63. If soccer <u>is</u> not as popular as it is today, people would watch other sports on TV.	2 2 2 2 2 2 Subtotal = 12	n/a		(Part 1) 2. If I <u>know</u> the answer, I would tell you. (Part 2) Grammar point: **2nd conditional (*if*-clause)** 2. **Rule:** (Sentence with 2 clauses required)	1	2	1 Subtotal = 4
3rd person –s in the simple present tense	10. A bad teacher **makes** lessons interesting and **cares** about students. 43. Eros Ramazzoti **sings** pop music. 64. Chicharito **plays** soccer with Chivas Rayadas. *4. Everyone loves comic books and <u>read</u> them. *28. A good student never <u>study</u> before an exam. *73. President Peña Nieto <u>speak</u> in public very often.	2 1 1 2 1 1 Subtotal = 8	1.Every morning Mr. Garcia **gets up** at 6:30 am. 2. **walks** to the store 3. **buys** a newspaper 4. He **returns** to this house 5. and **reads** the newspaper. 6. Then, if he **feels** like it, 7. he **goes** to work. 8. But often he **stays** at home 9. **drinks** a cup of coffee 10. and **sits** in the sun. 11. His wife **says** he should go to work. 12. But he always **smiles** 13. and **says** 14. "This ticket **belongs** to you", he said.	1 1 1 1 1 1 1 1 1 1 1 1 1 1 Subtotal =14	(Part 1) 3. Sara <u>cook</u> every day. (Part 2) Grammar point: **3rd person –*s* in the simple present tense** 3. **Rule:**	1	2	1 Subtotal = 4

Compara_ tive adjectives	1. Mexico is **greener** and **more beautiful** than other countries. 35. Zacatecas is a **nicer** place to visit than Leon. 53. Basketball is **more popular** than soccer around the world. *13. It is more harder to learn Japanese than to learn English. *65. Vicente Fernandez is more rich than Emmanuel. *75. Learning English is more easier than learning French.	2 1 1 1 1 1 Subtotal = 7	1. Some days are **hotter** than others. 2. One ticket was **older** 3. and **more** valuable than the other one. 4. A woman **younger** than him opened the door. 5. An **older** man came to the door. 6. I want to thank you for being **more** honest than most people I know.	1 1 1 1 1 1 Subtotal = 6	(Part 1) 4.This car is more cheaper than mine. 5.Mike is more tall than Joe. (Part 2) Grammar point: **Comparative Adjs.** 4. **Rule:** 5. **Rule:**	.5 .5	1 1	.5 .5 Subtotal = 4
Infinitives and gerunds (as verb comple_ ments)	2. Mexican people want **to keep** their country clean and green. 29. Children don't enjoy **going** to the beach in the summer. 66. Most famous singers started **singing** at a young age. *24. People in love usually want get married as soon as possible. *44. Students can't avoid to ask about exams. *54. Most young people enjoy to listen to music.	1 1 1 1 1 1 Subtotal =6	1. I want **to take** life easy. 2. I want **to enjoy** myself. 3. Every day I enjoy **reading** the… 4. I also like **sitting** in the sun. 5. I want **to thank** you for being more… 6. I want **to give** some money to you. 7. Would you mind **taking** five million… 8. He no longer needed **to work**. 9. In fact he and his wife enjoyed **spending** the money on…	1 1 1 1 1 1 1 1 1 Subtotal = 9	(Part 1) 6. The boys want buy a new car. 7. They finished to build the house. (Part 2) Grammar point: **Infinitives and gerunds (as verb complements)** 6. **Rule:** 7. **Rule:**	.5 .5	1 1	.5 .5 Subtotal =4
Indefinite article	20. Parents have **a** responsibility to care for their children. 67. Every city has **an** interesting museum to visit. 74. Music is **a** hard subject to study. *22. Every child needs good father. *70. You need to go to the library to buy the good book. *78. Paris is the good place to visit during vacations.	1 1 1 1 1 1 Subtotal =6	1. Every morning Mr. Garcia gets up at 6:30 am, walks to the store and buys **a** newspaper. 2. But often he stays at home, drinks **a** cup of coffee and sits in the sun. 3. Mr. Garcia found **a** wallet. 4. **A** woman younger than him opened the door. 5. Just **a** minute. 6. **An** older man came to the door.	1 1 1 1 1 1 1 Subtotal	(Part 1) 8. She bought the new house. (Part 2) Grammar point: **Indefinite article** 8. **Rule:**	1	2	1 Subtotal =4

			7. He thought it was **a** dream!	= 7				
Modal verbs + verb	6. People can **win** a lot of money in a casino. 40. Teachers must **prepare** their classes before they give a lesson. 55. Soccer players must **warm up** before the game starts. *11. Not everyone can to learn a second language. *51. People must to reserve a hotel before going to the beach. *68. Students should to study hard if they want to get a good grade.	1 1 1 1 1 1 Subtotal = 6	1. His wife says he should **go** to work. 2. He couldn't **believe** it. 3. She told him, "You must **return** the ticket to the owner." 4. Can I **see** Mr. Romo? 5. Mr. Romo couldn't **believe** that he had won the lottery. 6. Would you **mind** taking five million dollars?	1 1 1 1 1 1 Subtotal = 6	(Part 1) 9. I must to go to work. (Part 2) Grammar point: **Modal verbs + verb (simple form)** 9. **Rule:**	1	2	1 Subtotal =4
Many vs. much	19. People need **many** skills to learn English. 26. **Many** people study at university level today. 39. English teachers don't have **much** work. *30. Poor people need many money to travel. *56. Much students know Mexican history. *69. High school students don't need many time to study.	1 1 1 1 1 1 Subtotal = 6	n/a		(Part 1) 10. I have many money. 11. I didn't see much people at school today. (Part 2) Grammar point: **Many vs. much** 10. **Rule:** 11. **Rule:**	.5 .5	1 1	.5 .5 Subtotal = 4
Yes/no questions	25. It is difficult to ask, 'Do you really **love** me?' 45. It is an interesting question to ask, 'Does a teacher **want** a better job?' 57. It is a silly question to ask, 'Does a child **need** education?' *23. It is a silly question to ask 'Do a woman need to marry?' *38. It is a silly question to ask	1 1 1 1 1 1	n/a		(Part 1) 12. Do Pedro work late? 13. Did they took the book? (Part 2) Grammar point: **Yes/no questions** 12. **Rule:**	.5 .5	1 1	1

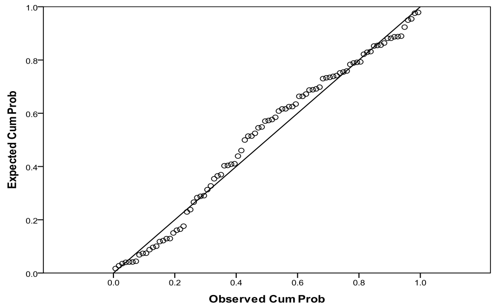
	'Does a student needs to study?' *50. It is an interesting question to ask 'Do you likes living in Aguascalientes?	Subtotal = 6						Total = 4
Plural of nouns	31. University **students** don't have many **books**. 46. It is not a good idea for **teachers** to punish **students**. 76. Bus **drivers** don't work more than 8 **hours** every day. *16. Young boys like fast car. *21. Rich people have two or three house. *58. A medicine student needs to study for five year to become a doctor.	2 2 2 2 1 1 Subtotal =10	1. Some **days** are hotter than **others**. 2. It contained some **bills** and **coins**, some **credit cards** and two **lottery tickets**. 3. Mr. Garcia checked the **numbers** of the **lottery tickets** in the newspaper. 4. It was worth 50 million **dollars**. 5. Would you mind taking five million **dollars**. 6. Mr. Garcia accepted the five million **dollars**. 7. In fact he and his wife enjoyed spending the money on buying new **houses** and visiting new **places** around the world.	2 4 2 1 1 1 2 Subtotal =13	(Part 1) 14. Joe sold his two car to a friend. (Part 2) Grammar point: **Plural of nouns** 13. **Rule:**	1	2	1 Subtotal =4
Since/For	12. To speak English well you must study **for** one year. 32. People have played sports **for** many years. 47. President Peña Nieto has been in the presidency **since** 1994. *8. People have used computers since many years. *59. Chicharito has played with Manchester United for 2006. *77. Vicente Fernandez has been the most popular 'rancheras' singer since three decades.	1 1 1 1 1 1 Subtotal = 6	n/a		(Part 1) 15. I have been here for 9 o'clock. 16. Teachers have used computers since two decades. (Part 2) Grammar point: **Since/For** 14. **Rule:** 15. **Rule:**	.5 .5	1 1	.5 .5 Subtotal =4
Direct and	7. People should report a car	1	1. He showed the ticket **to** his	1	(Part 1)			

indirect objects (Dative alternation)	accident **to** the police. 33. Students should give feedback **to** teachers. 72. Parents should give a good education **to** their children. *3. People should report stolen money <u>the police</u>. *49. You should give a present <u>your mother</u> on her birthday. *60. The government should give financial aid <u>poor people</u>.	1 1 1 1 1 Subtotal = 6	wife. 2. You must return the ticket **to** the owner. 3. Mr. Garcia showed the newspaper **to** the man. 4. He also gave the ticket **to** the man. 5. I want to give some money **to** you.	1 1 1 1 Subtotal = 5	17. The postman gave the letter <u>the woman</u>. (Part 2) Grammar point: **Direct and indirect objects (Dative alternation)** 16. **Rule:**	1	2	1 Subtotal =4
Relative clauses	5. The film that everyone likes **is** Star Wars. 48. The language that most people speak **is** French. 61. The two sports that most people watch **are** soccer and baseball. *9. The software that Bill Gates invented <u>it</u> changed the world. *34. The city that many people want to visit <u>it</u> is Paris. *71. The two political parties that the United States has <u>they</u> are the Republicans and the Democrats.	1 1 1 1 1 1 Subtotal = 6 Total = 93 points	n/a	Total: 73 points	(Part 1) 18. The car that my father bought <u>it</u> is new. (Part 2) Grammar point: **Relative clauses** 18. **Rule:**	1	2	1 Subtotal =4 Total: 52 points

Appendix K. Assumptions of multiple regression for explicit knowledge of difficult grammar points for the cohort of student participants as a whole

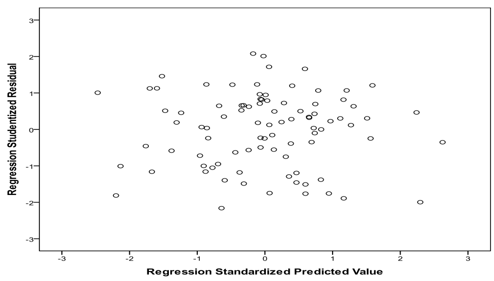
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-2.12, 2.03) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.087) which is below 1.0 and Mahalanobis distance (11.14) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the difficult grammar point scores on the MLK measure



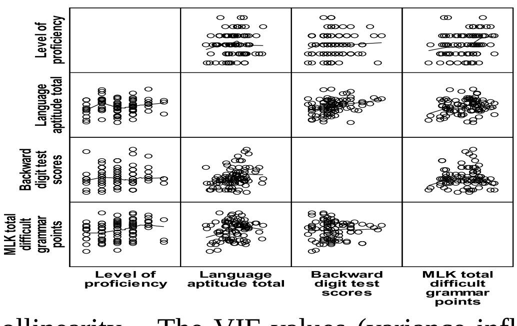
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.

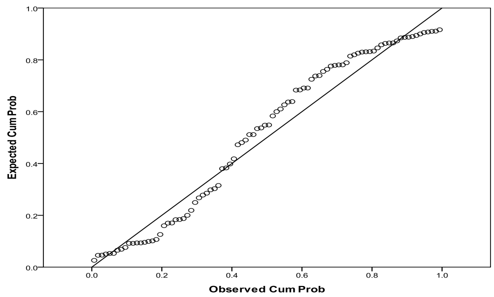


Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.13(language aptitude), which are under 5 and show no multicollinearity between variables.

Appendix L. Assumptions of multiple regression for explicit knowledge of easy grammar points for the cohort of student participants as a whole

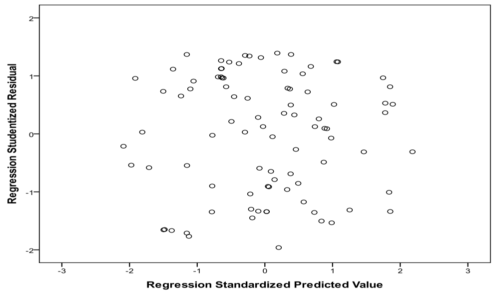
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-1.94, 1.38) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.076) which is below 1.0 and Mahalanobis distance (11.14) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the scores on EI measure.



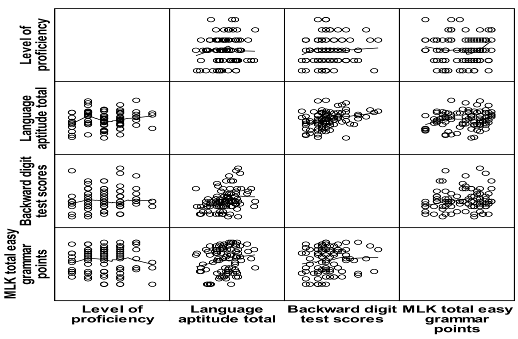
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.

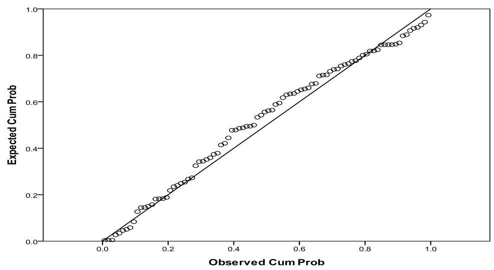


Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.13(language aptitude), which are under 5 and show no multicollinearity between variables.

Appendix M. Assumptions of multiple regression for implicit knowledge of difficult grammar points for the cohort of student participants as a whole

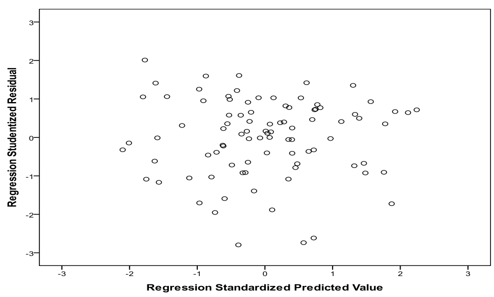
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-2.71, 1.93) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.12) which is below 1.0 and Mahalanobis distance (11.13) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the difficult grammar points scores on the EI/ON combined



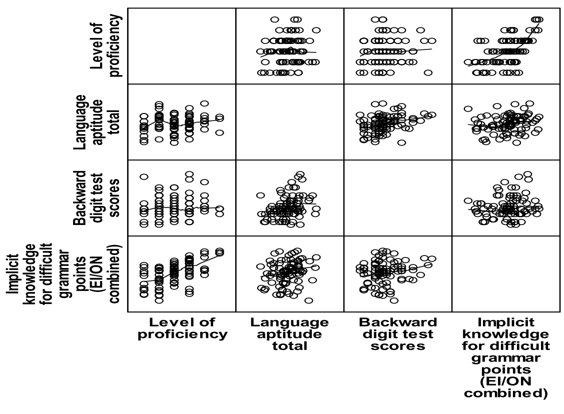
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.

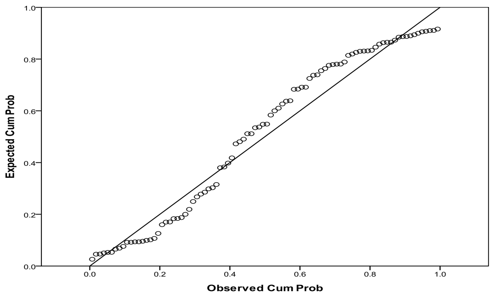


Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.13(language aptitude), which are under 5 and show no multicollinearity between variables.

Appendix N. Assumptions of multiple regression for implicit knowledge of easy grammar points for the cohort of student participants as a whole

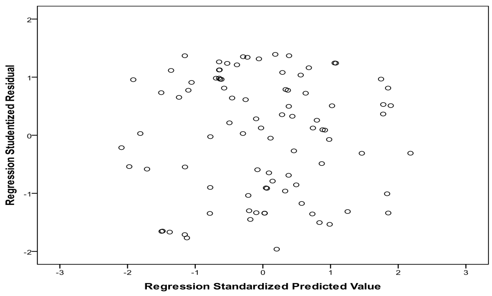
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-1.94, 1.38) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.076) which is below 1.0 and Mahalanobis distance (11.14) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the scores on EI measure.



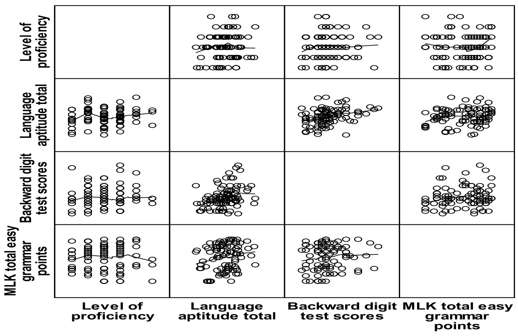
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.



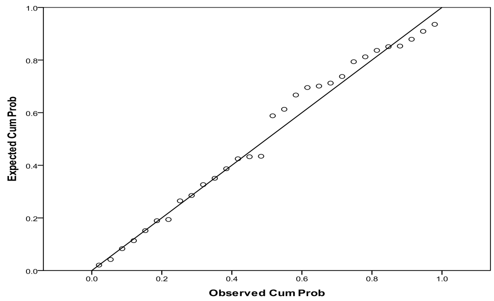
Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.13(language aptitude), which are under 5 and show no multicollinearity between variables.

Appendix O. Assumptions of multiple regression for implicit knowledge of difficult grammar points by group level

Level 5

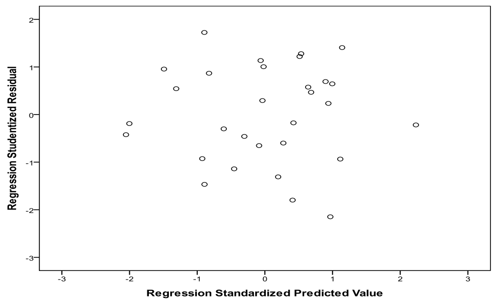
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-2.04, 1.52) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.21) which is below 1.0 and Mahalanobis distance (8.84) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the difficult grammar points scores on the MLK test



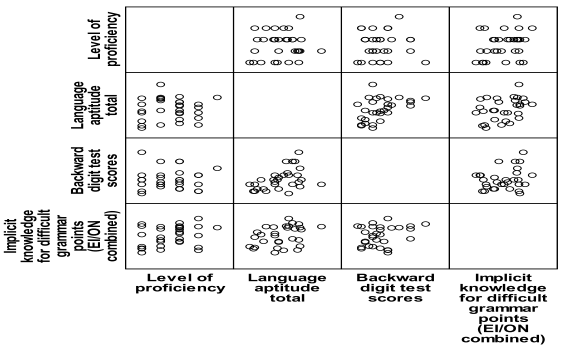
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.

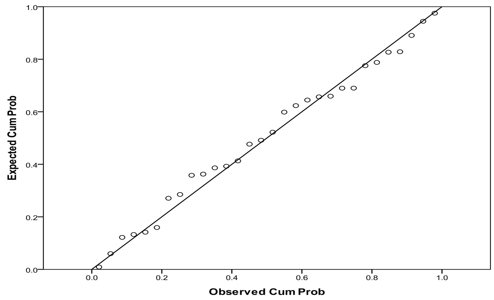


Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.27 (language aptitude), which are under 5 and show no multicollinearity between variables.

Level 7

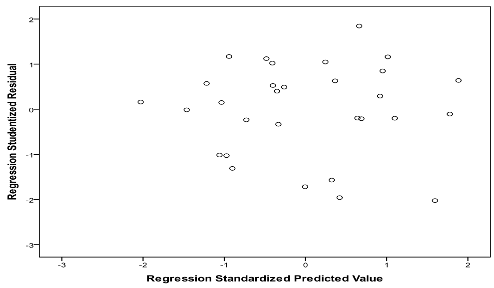
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-2.37, 1.96) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.21) which is below 1.0 and Mahalanobis distance (8.84) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the difficult grammar points scores on the MLK test



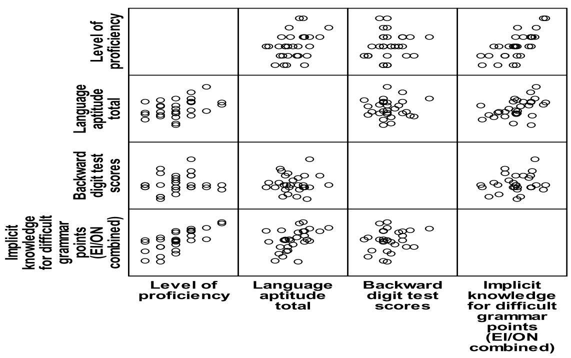
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.

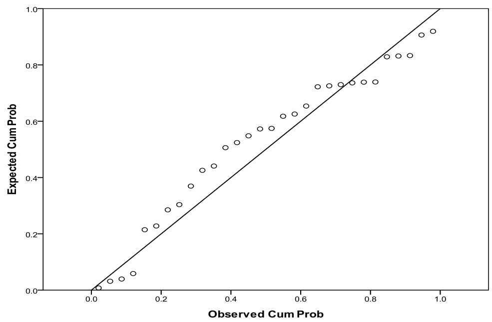


Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.18 (language aptitude), which are under 5 and show no multicollinearity between variables.

Level 9

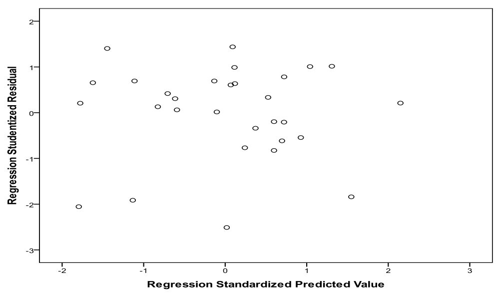
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-2.43, 1.40) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.32) which is below 1.0 and Mahalanobis distance (12.98) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the difficult grammar points scores on the MLK test



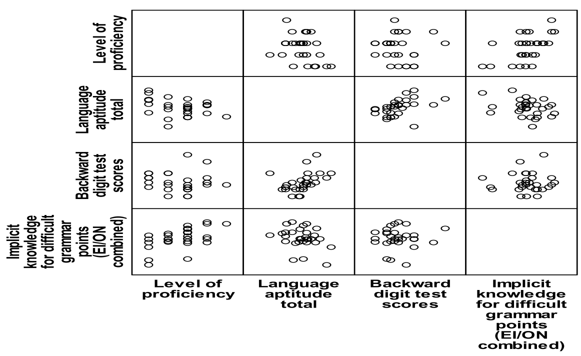
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.



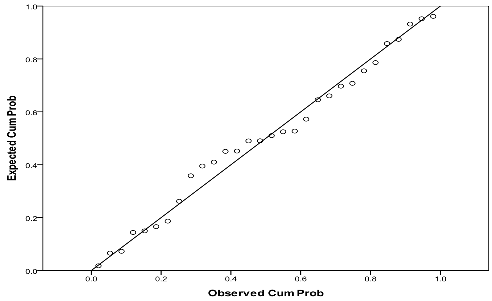
Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.37 (language aptitude), which are under 5 and show no multicollinearity between variables.

Appendix P. Assumptions of multiple regression for implicit knowledge of easy grammar points by group level

Level 5

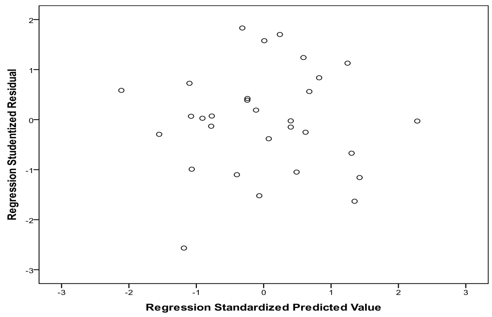
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-2.10, 1.76) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.82) which is below 1.0 and Mahalanobis distance (8.84) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the difficult grammar points scores on the MLK test



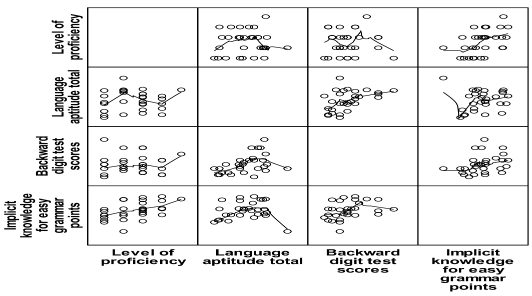
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.

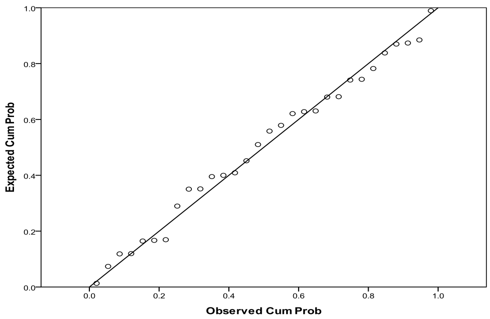


Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.27 (language aptitude), which are under 5 and show no multicollinearity between variables.

Level 7

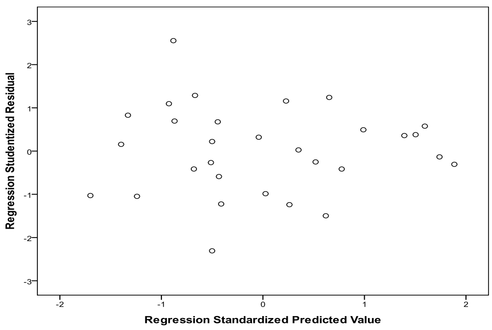
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-2.22, 2.31) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.37) which is below 1.0 and Mahalanobis distance (8.90) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the difficult grammar points scores on the MLK test



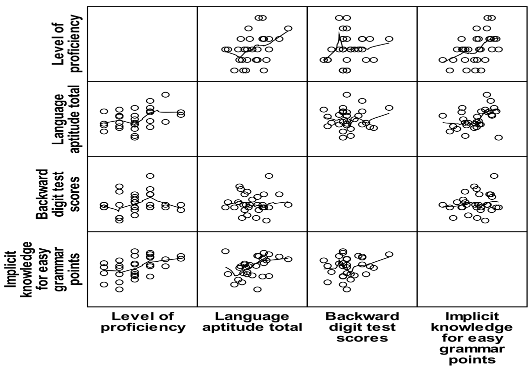
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.

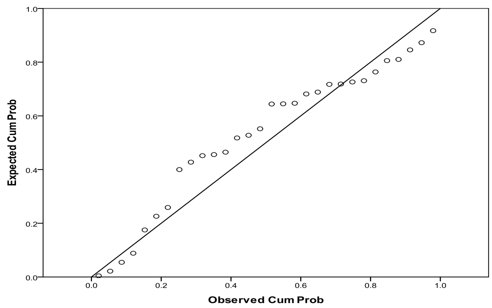


Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.18 (language aptitude), which are under 5 and show no multicollinearity between variables.

Level 9

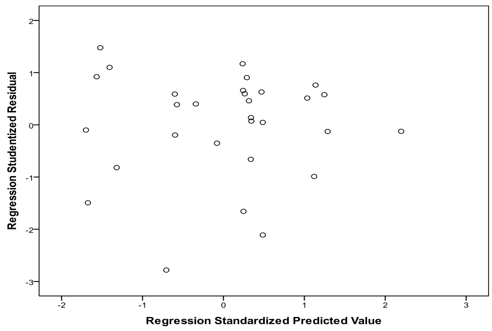
a) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-2.63, 1.29) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.22) which is below 1.0 and Mahalanobis distance (12.98) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the difficult grammar points scores on the MLK test



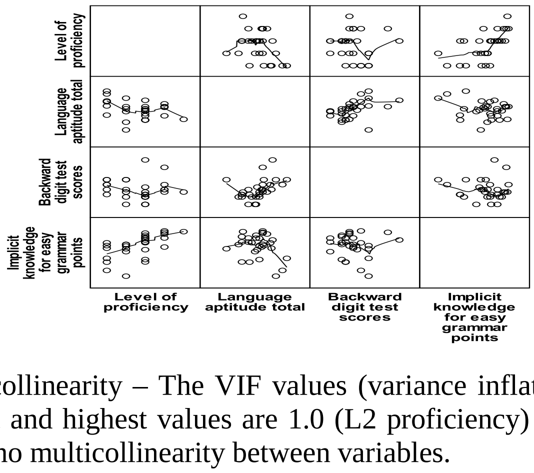
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.

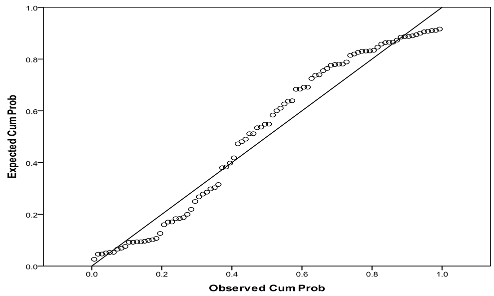


Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.38 (language aptitude), which are under 5 and show no multicollinearity between variables.

Appendix Q. Assumptions of multiple regression between sub-components of the LLAMA tests and explicit knowledge of difficult grammar points for the cohort of student participants as a whole

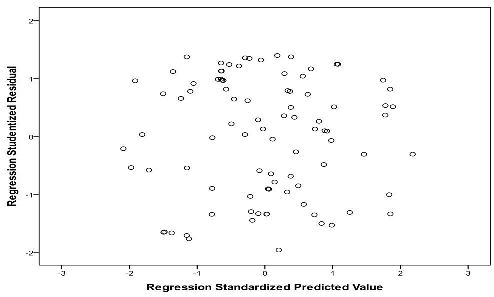
b) Normal distribution of data (distribution of the residuals) – The points in the P-P plot follow a linear distribution, the values for the standardized residuals (-1.94, 1.38) are not above 3.0 or below -3.0 so this test shows no outliers as indicated by the maximum values of Cook's distance (.076) which is below 1.0 and Mahalanobis distance (11.14) which is below 15.0 (see Figure 1).

Figure 1. P-P plot for diagnosing normal distribution of the scores on EI measure.



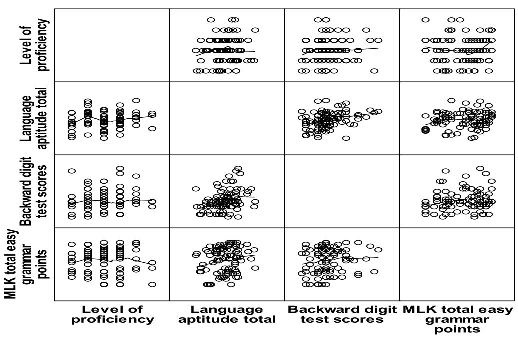
Homogeneity of variances – The shape of the scatterplot shows that the residuals are independent of the scores in the predictor variable, that is, the residuals of the outcome variable do not systematically increase or decrease as the scores in the predictor variable increase (see Figure 2).

Figure 2. Plot of studentized residuals crossed with fitted values



Linearity – The multiple scatterplot shows a reasonably linear relationship between the response variable (MLK difficult grammar points) and the predictor variables (L2 proficiency, language aptitude, and working memory) (see Figure 3).

Figure 3. Multiple scatterplot for predictor variables.



Multicollinearity – The VIF values (variance inflation factor) in the Coefficients table indicate that the lowest and highest values are 1.0 (L2 proficiency) and 1.13(language aptitude), which are under 5 and show no multicollinearity between variables.